DOCUMENT RESUME

ED 302 241 IR 052 572

TITLE	Database Search Strategies & Tips. Reprints from the Best of "ONLINE" and "DATABASE."
INSTITUTION	Online, Inc., Weston, CT.
PUB DATE	88
NOTE	101p.; For other volumes in this series, see IR 052 569-571.
AVAILABLE FROM	ONLINE, Inc., 11 Tannery Lane, Weston, CT 06883.
PUB TYPE	Collected Works - General (020) -- Journal Articles (080) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Libraries; Access to Information; Computer Graphics; Computer Software; *Databases; Expenditures; Higher Education; *Information Seeking; *Online Searching; *Search Strategies; *Technological Advancement; Telecommunications; *Users (Information)
IDENTIFIERS	Spreadsheets

ABSTRACT

Reprints of 17 articles presenting strategies and tips for searching databases online appear in this collection, which is one in a series of volumes of reprints from "ONLINE" and "DATABASE" magazines. Edited for information professionals who use electronically distributed databases, these articles address such topics as: (1) searching full-text periodicals; (2) how to use information search tactics online; (3) false drops; (4) graphing online searches with Lotus 1-2-3; (5) online and offline printing; and (6) tailoring search strategies for highly specialized topics. The articles and columns appearing in this volume have been selected from "ONLINE" and "DATABASE" issues from 1985 through early 1988, with the majority in the 1986-1987 period. (CGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED302241

DATABASE SEARCH STRATEGIES & TIPS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

database
ONLINE
online

BEST COPY AVAILABLE

REPRINTS FROM THE BEST OF ONLINE® database®

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jeffrey Pemberton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR052572

ONLINE INTERNATIONAL COMMAND CHART

A Two-Volume Set With MANY Professional Online Systems . . . Plus End-User Systems And CD-ROMs

Save Connect Time With Quick Reference To 42 Retrieval Commands On 31 Systems

The *Online International Command Chart* is a command-by-command comparison of the 42 major database searching commands. It stands next to your terminal or micro, ready to remind you of a command that you might be unsure of – for pre-search planning. . .*or while you're online*. It speeds your learning of new systems. . .and helps to maintain proficiency on systems that you don't use every day. How? By starting with a command that you're familiar with on your "home base" system – then moving your eye directly to the same command on the less used system.

The 1987/88 Edition Includes. . .What's New In The New Edition?

- Additional major professional online systems: WILSONLINE, Dow Jones, STN, NewsNet, DataTimes, WESTLAW, and reference commands for OCLC.
- Major CD-ROM retrieval systems: DIALOG OnDisc, Reference Technology and SilverPlatter are new this edition.
- Major End-user systems: CompuServe, DIALOG Business Connection, Knowledge Index, BRS Colleague and BRS/After Dark are new.
- Updated commands. Example: the DIALOG section now includes DIALMAIL, online editing features, user-defined formats, KWIC and Highlight commands, new DIALORDER and preformatted report commands.

Commands Prepared By Documentation Experts At The Online And CD-ROM Services

In almost all cases, commands were written (and proofread) by the specialists who prepare the official training materials. And, for the benefit of Command Chart users, they have made entries conform to uniform conventions, such as character spacing and term entry format.

Following are the systems included. New systems in the 1987/88 edition are marked with an asterisk.

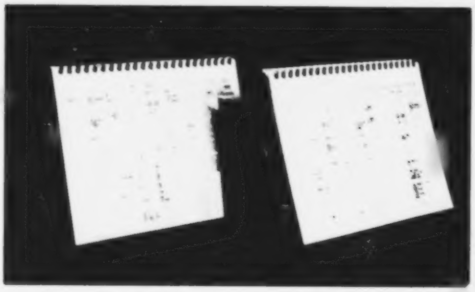
AMERICAN. . .DIALOG, BRS, WILSONLINE,* VU/TEXT, Dow Jones,* NewsNet,* DataTimes,* ORBIT, NLM, STN,* WESTLAW,* LEXIS/NEXIS, plus reference commands for RLIN and OCLC.*

CANADIAN. . .CAN/OLE, Info Globe, QL Systems

EUROPEAN. . .ESA-IRS QUEST, ESA-IRS Common Command Language, DIMDI,* Data-Star, Pergamon InfoLine, QUESTEL Plus

END-USER SYSTEMS. . .DIALOG Business Connection,* Knowledge Index,* BRS Colleague,* BRS/After Dark,* CompuServe*

CD-ROMs. . .DIALOG OnDisc,* Reference Technology,* SilverPlatter*



Forty-two major and minor system commands are listed, ranging from Logon to Logoff. Examples: FILE SELECTION, SEARCH COMMAND, SUBJECT TERM ENTRY, TEXT WORD ENTRY, COMBINING TERMS, TRUNCATION, PROXIMITY CONNECTORS, RESTRICTING SEARCH RESULTS, ONLINE PRINTING, DOCUMENT DELIVERY. . .plus 32 more!

Pays For Itself Over And Over In More Efficient Searching

Standing next to your terminal, the *Online International Command Chart* is a ready reference that serves the same function as the handy "cheat sheets" that come with many software programs. . .a fast way to remind yourself of a key command without thumbing through a mountain of documentation. The format, itself, is designed for heavy use at the terminal. . .large type, heavy, durable paper in a flip chart with O-ring binding. A Velcro strap keeps it standing. Size: 9 x 12 inches. Pages: 76 per volume. Total number of systems: 31.

Order Now

The *Online International Command Chart* 1987-88 Edition costs $88.00. Additional charts are $70.00 each when included in the same order to the same mailing address as the first ordered at the regular price. Simply fill out the form below to order and mail to: Online, Inc., Dept CC, 11 Tannery Lane, Weston, CT 06883, telephone (203) 227-8466. To order, call **TOLL-FREE 800-248-8466**

Name & Title ____________________

Organization ____________________

Address ____________________

City/ST/Zip ____________ Country ____________

Phone ____________________

Please send me sets of the *Online International Command Chart* 1987-88 Edition. Shipping charges will be added to foreign orders.

☐ Please invoice Charge to: VISA ☐ MasterCard ☐ AMEX ☐ Acct.# ____________

Exp/date ____________ ☐ Check enclosed

SEARCH STRATEGY & TIPS

TABLE OF CONTENTS

ONESEARCH: HOW AND WHEN TO USE IT by Ruth Pagell 1

SEARCHING FULL-TEXT PERIODICALS: HOW FULL IS FULL? by Ruth Pagell 9

SEARCHING IAC'S FULL-TEXT FILES: IT'S AWFULLY CONFUSING by Ruth Pagell 13

HOW TO USE INFORMATION SEARCH TACTICS ONLINE by Marcia J. Bates 21

FALSE DROPS How They Arise...How To Avoid Them by Elaine Wagner 29

ZERO RESULT SEARCHES...HOW TO MINIMIZE THEM by Steve Atkinson 33

GRAPHING ONLINE SEARCHES WITH LOTUS 1-2-3 by Olle Persson 41

A PRIMER ON CROSSFILE SEARCHING IN THE INFORMATION ACCESS COMPANY DATABASES by Judith Sovner-Ribbler 44

PASSWORD PROTECTION: Some Dos and Don'ts by M. D. Parker 51

"MEATBALL SEARCHING" The Adversarial Approach To Online Information Retrieval 55

GATEWAYS: AN INTRODUCTION (THE FRIENDLY USER) by Lucinda D. Conger 63

ONLINE AND OFFLINE PRINTING...RELATIVE COSTS by Brian Williams 65

SEARCHWORKS: DOES IT REALLY WORK? by Bill Coons 69

Communications Costs For Database Access – Waiting Is Costly by Ralph Lee Scott 76

WHAT'S IN A NAME? USE OF NAMES AND TITLES IN SUBJECT SEARCHING by Anne B. Piternick 82

"Hedges"...The Sometimes Ignored Search Technique That Can Save A Lot of Time by Genevieve N. Schiffmann 89

Tailoring Search Strategy for a Highly Specialized Topic...Proof Once Again of the Necessity of Using File Documentation to the Fullest by Maureen Lambert and Fern Brody 91

SEARCH STRATEGY & TIPS

PREFACE

This volume is part of a six volume series of reprints from *ONLINE* and *DATABASE* magazines. Both are edited for information professionals who use databases distributed electronically. Most readers do their searching through central databanks, such as DIALOG. Most of the coverage in *ONLINE* and *DATABASE* reflects this fact. However, a certain amount of coverage is also devoted to other media, such as laserdisks. (full coverage of laserdisks is in a companion journal, *the Laserdisk Professional.*)

The following articles and columns have been selected from *ONLINE* and *DATABASE* issues from 1985 through early 1988, with the majority in the 1986-87 period.

Other titles in the *Best of ONLINE and DATABASE* series include:

- Library Microcomputing(two volumes)
- Business and Finance(two volumes)
- Science & Technology
- Life Sciences
- End-users & Public Access

We hope you find these reprints useful. If you have any comments or questions, please write or call me directly toll free. Don't be bashful; we're a small shop and I love to hear from readers.

Jeff Pemberton, President and Publisher
Online, Inc.
11 Tannery Lane, Weston, CT, 06883
800/248/8466 Note: subscription orders and other information on *ONLINE, DATABASE* and *the Laserdisk Profesional* can also be obtained through the 800 number.

ONESEARCH: HOW AND WHEN TO USE IT

by Ruth Pagell
Associate Librarian
Lippincott Library of the Wharton School
University of Pennsylvania

Editor's Note: To take a good look at DIALOG's new OneSearch capability, we looked for a super-searcher, one who could test the limits of OneSearch and acquaint DATABASE *readers with the pros and cons of using this new feature. When DIALOG reviewed the text of this article prior to publication, they were distressed that our author recommended OneSearch only to experienced searchers, and then that it be used only after careful pre-search planning. According to DIALOG, many searchers use OneSearch to search only the basic index, do not use field tags and controlled vocabulary, and thus do not encounter many of the problems experienced by a heavy searcher executing a precise or comprehensive search. Additionally, DIALOG feels that even single file searching at a high level requires careful planning, and that OneSearch was unnecessarily criticized in this regard.*

Multifile, cross-file, and/or cluster searching are capabilities being added to several online systems. DATABASE *will continue to cover and compare these capabilities, and welcomes comments or suggestions for future articles on the topic from searchers at different levels, or in different disciplines. —NG*

"How many times have you wished you could search more than one database without having to re-enter your search terms or execute a SearchSave. .?" the October 1987 *CHRONOLOG* asks. How many times have you wished that there were consistent field tags and field entries, cross-database vocabulary control, similar pricing across databases, and some way of merging search results and eliminating duplicates as you moved from one file to another? On October 1, 1987 DIALOG introduced OneSearch. As the name implies, one search can be performed across a variety of databases at one time. But is this new cross-file searching capability the boon we hoped it to be? Are we really getting anything more than we had before? DIALOG's OneSearch answers our first wish while leaving us to cope with all the other inconsistencies involved in moving from one DIALOG database to another.

Cross-file searching is not unique to DIALOG. You can select one file or a group of files on LEXIS/NEXIS. More than one WILSONLINE index can be searched at once. Dow Jones introduced multifile searching in **//TEXT** about a month before DIALOG's official OneSearch announcement. In late October, VU/TEXT activated its global search software. DIALOG's Business Connection (DBC), available for over a year, is a menu-driven gateway to multiple business files. ESA-IRS now allows "cluster" file searching.

In this article, I examine cross-file searching on DIALOG's OneSearch. I describe OneSearch, consider situations where OneSearch is appropriate, discuss some of the pitfalls of searching across DIALOG databases, and offer some final opinions.

METHODOLOGY

I intended to perform different types of searches both with and without OneSearch in order to compare time, cost, results, and ease of use. However, since it proved difficult to hold all the variables constant, I will focus on search techniques and search strategies rather than on cost/time comparisons.

Thanks to the generosity of DIALOG, I spent many hours on OneSearch, performing

►
1

THE
Online
Chronicle

File 170 on DIALOG
From the Publisher of
ONLINE and *DATABASE*

The ONLINE CHRONICLE — a full-text database of advance news and editorials from *ONLINE* and *DATABASE* magazines. Each biweekly issue is edited to keep information users informed on North American/European News, Database Updates, Hardware/Software Developments, Search Aids & Publications, Meetings & Conferences, Book Reviews and JOBLINE. . .free listings of Positions Available and Positions Sought for employers and employees alike.

To see the most recent Table of Contents, logon, **b 170** and **Select TC;T1/5**. A specific article may be selected by individual accession number — The complete file can be searched retrospectively by keywords, free-term or accession number — Ongoing articles are updated each issue and retain the same accession number such as: Jobline Table of Contents: AN = 001696

For more information on how to search the ONLINE CHRONICLE, or to list your Position Available/Position Sought on JOBLINE, contact June Thompson, Editor or Tasha Heinrichs, Jobline Editor, Online, Inc., 11 Tannery Lane, Weston, CT 06883; 203/227-8466; (OnTyme-CLASS.ONLINE; The Source-TCU202; BRS-TFHN; ALANET-ALA0795)

**THE ONLINE CHRONICLE — File 170 on DIALOG
For Most Recent Table of Contents —
Select TC;T1/5**

searches and cross-checking print documentation with online displays. Searches were performed between October 4 and November 28, 1987.

ONESEARCH – THE BASICS

Logging on: To use OneSearch at any **?** prompt, use the BEGIN command and select a DIALINDEX category or create a set made up of files of your choosing (Figure 1). OneSearch can handle up to 20 files at once.

You can access OneSearch immediately upon logon or at any point during a search session. It is available on all files and will also work with an instructional password (in files you can normally access with that password).

FIGURE 1

LOGGING ON TO ONESEARCH

```
?b1,61,202,275

        12oct87 05:54:16 User999999
              $.027   0.009 Hrs File1
   $0.27      Estimated cost File1
   $0.07      Dialnet
   $0.34      Estimated total session cost   0.009 Hrs.

System:OS – DIALOG OneSearch

   File   1:  ERIC – 66-87/SEP

   File  61:  LISA – 69-87 (8709)

   File 202:  INFORMATION SCIENCE ABSTRACTS 66-87/JAN
              (C. IFI?PLENUM DATA CO. 1987)

   File 275:  COMPUTER DATABASE 83-87/ISSUE19
              (COPR.1987 INFORMATION ACCESS COMPANY)
```

Choosing multiple files or a DIALINDEX category with the **BEGIN** command will initiate OneSearch.

FIGURE 2

ENTERING SEARCH STRATEGY

```
?ss expert(w)system? ?/de and scheduling

 Processing
    S1      6605   EXPERT/DE
    S2    386016   SYSTEM? /DE
    S3      6477   EXPERT/DE(W)SYSTEM? ?/DE
    S4     18520   SCHEDULING
    S5       174   EXPERT(W)SYSTEM? ?/DE AND SCHEDULING
```

Enter the search strategy in OneSearch as usual. The results will be aggregates of the files searched.

Searching: **SELECT** statements are input in OneSearch as they would be in any DIALOG search. The system searches across files and one result is displayed (Figure 2).

The number of hits per database can be displayed by issuing the command, **SET DETAIL ON**, either before you enter your search or at a point where you wish to **DISPLAY SETS** (Figure 3).

The **EXPAND** command can also be used in

OneSearch. If the **SET DETAIL** command is not on, the expanded output looks like usual DIALOG output. With **SET DETAIL ON** a breakdown of hits per database is displayed, though only one "E" reference per term is assigned (Figure 4).

If you go directly into OneSearch at logon, you might wish to follow this procedure:

1. Enter your search strategy
2. If you have too few/too many hits, SET DETAIL ON
3. DISPLAY SETS
4. Revise your search
5. SET DETAIL OFF

Output: Citations are displayed database-by-database, in the order in which you listed your files. All of the citations from the first database are shown, as they would normally appear. Then all of the citations from the second database are displayed, etc. You may choose any of the standard formats or try user-defined formats to minimize the differences in formats among databases. Reports and MAPping are not available.

Search results can be displayed from an individual database with the **FROM** option:

```
T30/3/ALL FROM 1
```

or you can request items from different databases by using ranges: **30/3/1-3, 17-26,33-34**.

Sorting: The **FROM** option can be used to sort in individual databases. A separate set is created for each sort. You can display one set at a time, or combine the sorted output into a single set.

FIGURE 3

DETAILED SEARCH RESULTS

To display the number of hits in each database, **SET DETAIL ON** before beginning your search.

```
?set detail on

DETAIL set on

?ss expert(s)system? ?/de and scheduling

 8: COMPENDEX - 70-87/SEP
             971   EXPERT/DE
          116956   SYSTEM? ?/DE
             971   EXPERT/DE(W)SYSTEM? ?/DE

 13: INSPEC - 77-87/ISS22
Processing
            4977   EXPERT/DE
          158649   SYSTEM? ?/DE
            4977   EXPERT/DE(W)SYSTEM? ?/DE
TOTAL: Files 8,13,63 and . .
   S1       6605   EXPERT/DE
   S2     386016   SYSTEM? ?/DE
   S3       6477   EXPERT(W)SYSTEM? ?/DE
```

Or use **SET DETAIL ON** to review your search results:

```
?ds

        Set    File      Items     Description
                  8        971
                 13       4977
                 63         74
                 15        551
                 16         32
        S1                6605     EXPERT/DE
                  8     116956
                 13     158649
                 63      30528
                 15      27479
                 16      52404
        S2              386016     SYSTEM? ?/DE
```

FIGURE 4

EXPANDING

```
?b15,75,49,139
          26nov87   06:33:45   User99999

        Set   Items   Description

?e jn=journal of economic 1

Ref      Items   Index-term

E1       4503    JN=JOURNAL OF ECONOMIC HISTORY
E2       2895    JN=JOURNAL OF ECONOMIC ISSUES
E3         *0    JN=JOURNAL OF ECONOMIC L
E4       3300    JN=JOURNAL OF ECONOMIC LITERATURE
E5        290    JN=JOURNAL OF ECONOMIC PSYCHOLOGY
```

EXPAND*ing with* **SET DETAIL** *not on.*

```
?set detail on
DETAIL set on
?e jn=journal of economic l

Ref     File     Items    Total    Index-term

E3      --                    0    *JN=JOURNAL OF ECONOMIC L
          75        39                JN=JOURNAL OF ECONOMIC LITERATURE
           7      2969                JN=JOURNAL OF ECONOMIC LITERATURE
         139       292                JN=JOURNAL OF ECONOMIC LITERATURE
E4      --                 3300     JN=JOURNAL OF ECONOMIC LITERATURE
```

EXPAND*ing with* **SET DETAIL ON**, *gives the same "E" numbers and a breakdown by database.* ►

Results are *not* merged, only the order of the records *within* each database will change.

THE ONLINE PRICE OF ONESEARCH

DIALOG has developed a pricing algorithm that calculates cost based on time spent in each database. The actual search and display time per database is charged at that database's rate. You pay the same citation charges as in a regular DIALOG search. Any additional online time, while you are typing, thinking, sorting, expanding, or caught in communications downtime is charged back across the databases, prorated on the actual time spent searching and displaying in each file (Figure 5).

FIGURE 5

SESSION COST DISPLAY

Costs are based on time the system spends searching each database, and are prorated for "thinking" time. Note the use here of the **FROM EACH** command, which can choose which records are **TYPE**d from which database.

```
?ss affinity (10w) (credit()card?)

    S1      603      AFFINITY
    S2    55684      CREDIT
    S3    34695      CARD?
    S4       35      AFFINITY (10W) (CREDIT()CARD?)

t4/3/1 FROM EACH

4/3/1     (Item 1 from file: 15)
.
.
.
4/3/5     (Item 1 from file: 16)
.
.
.
4/3/22     (Item 1 from file: 148)
.
.
.
?logoff

            14dec87 10:16:43 User xxxxxx
                $0.58     0.006 Hrs File15
                    $0.34 1 Types in Format 3
                $0.34 1 Types
        $0.92 Estimated cost File15
                $1.26   0.010 Hrs File16
                    $0.63 1 Types in Format 3
                $0.63 1 Types
        $1.89 Estimated cost File16
                $2.44   0.029 Hrs File148
                    $0.00 1 Types in Format 4
                $0.00 1 Types
        $2.44   Estimated cost File148
                OneSearch, 3 files, 0.046 Hrs FileOS
        $0.37 Dialnet
        $5.62 Estimated cost this search
```

WHEN TO USE ONESEARCH

Obviously, many searches do not require cross-file searching. The nature of some material and the very specific protocols of some databases do not lend themselves to this technique even when more than one database is required. Where files have been divided into backfiles and current files, like the patent files or **SCISEARCH**, OneSearch is a natural. Searching all the SCISEARCH files is useful for cited references. For a patron tracking trademarks, you will want to use the federal and state trademark databases together.

Most searches, however, are not that clear-cut. Whether or not to use OneSearch depends on your answers to the following questions:

- How are you searching now?
- How much information do you need?
- What is the record format?
- How do you wish to use the output?
- How experienced a searcher are you?

How Are You Searching Now?

The value of OneSearch to you depends, in part, on how you are presently preparing for and structuring your DIALOG searches. Just thinking about OneSearch is positive, because in order for you to evaluate whether to use OneSearch, you have to evaluate your present searching methods. Those of us who have been searching for many years have developed patterns that may not always be appropriate. If you are not currently searching multiple files, experimenting with OneSearch might encourage you to broaden your database scope.

There are many alternative search methods:

- Search in your one favorite database
- Search in one database; save search; re-execute
- Search DIALINDEX; save search; search individual databases
- Search DIALINDEX; save search; use OneSearch
- Search OneSearch directly

DIALOG recommends that you begin in DIALINDEX, then move into OneSearch.

How Much Information Do You Need?

Common sense dictates that if only a few records are needed, using more than one database is not necessary. If either a very narrowly-defined search or an exhaustive search is required, it may be inappropriate to use OneSearch. Because several databases are being accessed at one time, there will be more hits per search statement than you usually expect. There are certain limitations to OneSearch that affect both precision and recall.

What is the Record Format?

The type of record, bibliographic, full-text, numeric, or directory, is relevant in making a decision to use OneSearch. In general, searches across bibliographic databases lend themselves to OneSearch. If you like full-text searching on DIALOG, then try OneSearch. There is little advantage to using OneSearch for numeric or directory searches, since results cannot be merged, and the report format is not available.

Also, there is no advantage to OneSearch if the type of records and information differ from database to database. I tried to duplicate a Datext CDROM search on "Lockheed Corp." by constructing my own Datext category on OneSearch from **ABI/INFORM**, **PROMT**, **DISCLOSURE**, **Who's Who**, **INVESTEXT**, and **Media General**. Results showed this was not an appropriate mix for OneSearch.

How Do You Wish to Use the Output?

If you are using a bibliographic database package like ProCite, you cannot use OneSearch without major reconfiguration of the software. On the other hand, simple editing of OneSearch results are probably easier because there is only one display of search strategy. Numeric information cannot be downloaded in a format for use with spreadsheet or database software, since REPORT is not available.

How Experienced a Searcher Are You?

Know yourself. Knowledge of the content and structure of the databases is very important. Novices, casual searchers, and end-users should NOT be using OneSearch. Even an experienced and prepared searcher will find surprises. It is more difficult to pinpoint problems and revise strategy. If you are a searcher who spells out each step ahead of time and has difficulty with online revisions, then OneSearch is not for you.

CONSTRUCTING YOUR ONESEARCH

It is easier to search across databases on the other systems offering this capability than to do so on DIALOG, since in general, other systems have more consistency in database design, record type, and pricing. Searching with OneSearch requires a rethinking of search strategy construction. An up-to-date set of blue sheets is a prerequisite for going online.

Some of the pitfalls and problems involve:

- Field tag labels
- Field entries
- Descriptor terms
- Updating
- Displaying results

These are problems experienced searchers have always faced in moving from file to file, or in using DIALINDEX, but they are magnified in OneSearch.

Field Tag Labels: Familiarity with field labels is essential. DIALOG has been reloading databases to make field tags consistent. Still, there will always be differences among databases in number of fields, field labels, and meanings of field labels, due to independent database producers' file design. In some databases, the same label can actually stand for two fields (Table 1). CODEN is **CO=** in most files but **CO=** in ABI/INFORM. However **CO=** is both CODEN and COMPANY in **Management Contents**.

Field Entries: Even if the field tags are the same, entries in these fields vary. Authors' names and corporate source are two examples. Authors' names can vary within databases. Social SciSearch and SciSearch not only just use initials, but even the current reloaded version does not include a comma in the author's name. In truncating to accommodate all database variations, false hits are inevitable. For example, the corporate name for University of Pennsylvania is entered:

LISA (File 61)	**University of Pennsylvania**
Conference Papers (File 77)	**Univ Pennsylvania**
PsycINFO (File 11)	**U Pennsylvania**

The number of digits is not the same in numeric fields. Net sales (**SA=**) have 14 digits in DISCLOSURE, but only 12 in the Dun & Bradstreet and Trinet files.

Descriptors: Precise descriptor terms differ even among databases with similar subject areas. A free-text OneSearch for "crossfile searching" in the library literature turned up one article in ERIC (File 1), LISA (File 61), and Information Science Abstracts (File 202), entitled "A Cognitive View of Three Selected Online Search Facilities," by Peter Ingerson in the October, 1984 *Online Review*. While many similar words appeared in the descriptor fields for these three databases, the exact phrases varied (Table 2).

I recommend that each descriptor word be searched individually with **/DE** rather than entering a phrase.

Searching for company names and industry codes is tricky even for experienced business database users. A company with a distinctive, one-word name, like GODIVA, has relatively consistent entries among databases; a company with a name like W R GRACE may have variations even within the same database.

Use of the ticker symbol (**TS=**) for publicly traded companies or a D-U-N-S number (**DN=**) for public or private companies is one way of controlling for name variations in files that use either of these fields.

DIALOG Business Connection has a "Company Finder" feature which provides a cross-database index of company names. If you enter a part of a company name, i.e. GRACE, all com-

►

TABLE 1

FIELD TAGS IN BIBLIOGRAPHIC BUSINESS DATABASES

Field Label	Name	ABI/INFORM File 15	Mgmt. Contents File 75	PROMT File 16	T&I Index Files 148,648
AT	Article Type		**AT = Biography**		**AT = Biography**
AU	Author	**AU = Brod, Craig?**	**AU = Brod, Craig?**		**AU = Brod, C?**
CC	File dependent[1]	**CC = 7100**		**CC = 1USA**	**CC = 413 N.E.2D?**
CD	CODEN	**CD =**			
CN		**CN = Marketing**	**CN = A vs B**	**CN = United States**	
CO	CODEN		**CO =**		Limited
CO	Company Name	**CO = Lockheed?**	**CO = Lockheed?**	**CO = Lockneed?**	
DN	D-U-N-S Number	**DN =**		**DN =**	
DT	Document Type	**DT = Journal Paper**			
GC	Geographic Location		**GC = NMCN**		**GC = AEJA**
GN	Geographic Name		**GN = Canada**		
JN	Journal Name	**JN = Jrnl of. . .**	**JN = Journal of. . .**	**JN = Journal of . . .**	**JN =**
LA	Language	**LA = English**		**LA = NONENGLISH**	
NA	Named Person		**NA = Jobs,S?**		
PC	Product Code			**PC = 3573**	
PD	Publication Date	**PD = 8601**	**PD = 860222**	**PD = 8703**	**PD = 860415**
PN	Product Name		**PN = Microsoft?**	**PN = Microwave. . .**[2]	**PN = Microsoft**
PY	Publication Year	**PY = 1986**	**PY = 1986**	**PY = 1987**	**PY = 1986**
SC	SIC Code		**SC = 3573**		**SC = 3573**
SF	Special Feature		**SF = Photograph**	**SF = Table**	**SF = Map**
SF	Subfile		**SF = MC**		**SF = Magazine Index**
SO	Source(16);Statute(148)			**SO = Wear**	**SO = 11 U.S.C. 15**
TS	Ticker Symbol		**TS = X**	**TS = BUD**	**TS = BUD**
Sortable fields		**AU,CO,JN,PY**	**AU,JN,PY,TI**	**CC,CO,JN,PD,PY**	**AU,JN,PY,TI**

[1] Classification Code#15
Case Citation#148
Country Code#16

[2] Text description of **PC**

TABLE 2

DIFFERENCES IN CONTROLLED VOCABULARY IN FILES 1, 61, and 202

FIELD	ERIC, File 1	LISA, File 61	INFORMATION SCIENCE ABSTRACTS, File 202
DE	*Artificial Intelligence* *Cognitive Processes* *Cognitive Structures* *Information Processing* *Information Retrieval* *Information Systems* *Man Machine Systems* *Models* *Online Systems* *Search Strategy* *Subject Index Terms*	*Command Languages* *Man-Machine Interface* *User-System Interface* *Interface* *System-User Interface* *Information Storage and Retrieval* *Information Retrieval* *Subject Indexing* *On-line Information Retrieval* *Searching* *Computerized Information Storage and Retrieval*	*INFORMATION STORAGE AND RETRIEVAL—Searching, Search Strategy.*

panies with GRACE anywhere in their name in any of the DBC databases are listed.

Industry information, accessed by an industry name or code, is database-dependent. SIC codes present problems all their own which have been magnified this year with the introduction of revised codes by the U.S. government. Only Trinet (Files 531, 532) has adapted the revised codes at this time.

EXPAND: **EXPAND** helps identify alternative entries. But even **EXPAND** will not pick up alternative journal titles if you include ABI/INFORM in your file set. If you are looking for articles in *Journal of Economic Literature* and **EXPAND** across Files 15, 75, 139, and 49, you get a total of 3300 hits (Figure 4). Only by **EXPAND**ing with **SET DETAIL ON**, do you see that File 15 is not included. To retrieve the *Journal of Economic Literature* or any "Journal" in ABI/INFORM, the abbreviation "jrnl" is used:

```
?b15
?s jn=jrnl of economic literature
      S1     115     JN=JRNL OF ECONOMIC LITERATURE
```

Expanding an author's name, i.e., **e au=emery, k?** would not retrieve any entries from the SCISEARCH and Social SciSearch files, or from parts of **COMPENDEX**, due to the comma.

Searches requiring **EXPAND**s on several fields are better executed on a database by database basis.

Updating: Check an old or saved search you are planning to update very carefully before plugging it into OneSearch. The databases used might have different updates. On November 30, 1987 the following banners were displayed:

```
File 1:ERIC -66-87/NOV
File 202:INFORMATION SCIENCE ABSTRACTS 66-87/APR
File 61:LISA -69-87 (8710)
```

If you originally searched in single files, you may have used database-specific strategy. I wanted to update a search on in-house advertising, performed in May 1985, as an example. I planned to re-enter the search strategy and limit by update (**UD**). However, the updates varied, so this did not work well in OneSearch.

Displaying Results: Output formats may differ among databases. If using both full and free-text, use of KWIC is an alternative. However, not all full-text files have the KWIC option, as in Trade & Industry ASAP (648) and INVESTEXT (545).

User-defined formats are another alternative. If you feel you must use directory files in OneSearch, UDF is not operative in the D&B files.

Database producers set their own format content and rates, so searchers should check both before printing output from OneSearch. In bibliographic files, where the output formats are consistent, the cost for one typed full record (Format 5) can range from $.14 in ERIC to $.68 in ABI/INFORM.

DUPLICATE RECORDS—A RED HERRING

Immediate reaction to OneSearch, when it was announced at ONLINE '87 in Anaheim, was vocal concern over duplicate records. The system does not identify or eliminate duplicates. There is no way to merge records from different files.

In OneSearch, records can be sorted, database by database, combined, and then printed. Or they can be sorted and then printed by database. In a search for indexing certificates of deposit several databases including **American Banker** were searched. This file was displayed first. Then the journal *American Banker* was "NOTted out" of the sets from other databases to eliminate the duplicates.

There is currently no provision for identifying duplicates in DIALOG. Any weeding must be done manually. Knowledge of database coverage, predicted overlap, and online costs of typing records should serve as a guide to handling duplicates.

How are you weeding now? Do you:

- Not weed at all (often the easiest when downloading)?
- Print everything from your first database; then scan titles from the others?
- Begin in the least expensive file, and print out all hits there first?
- Print all titles; go offline to check and then go back online?

You can use any of these methods in OneSearch. One scenario is to use **SET DETAIL ON** to see which database has the most hits. Sort results by title or author, if possible. Take full records from the database with the most hits first, print out titles from the others and go offline to check for original records. If one database is much less expensive than the others, like ERIC, you might still prefer to do the search there first, get all the results, **SAVE TEMP**, and then search LISA and Information Science Abstracts together.

DIALOG is working on an enhancement that will merge sorted records and, ideally, flag duplicates. Sometimes duplicate abstracts are desirable. DIALOG can not delete a record, since the system cannot decide for the user which file's record to delete.

Both Dow Jones and VU/TEXT provide alternatives to a database by database output of results from cross-file searching. Dow Jones //TEXT does a chronological merge.

The VU/TEXT global search software is

► 7

capable of sorting by "rank order." You may choose to view your results based on the number of occurrences of the search term in a record.

For example, using VU/TEXT, a predefined group of databases, BUS, was selected to look for information about the investment banking firm of Hambrecht & Quist. The firm's name was limited to the headline and lead paragraphs fields. VU/TEXT alerts users to avoid keyword searching in global groupings of files, and suggests this method to limit the search.

Worrying about duplicate records on OneSearch relates to two important issues: cost and time; not only online cost and time, but your cost and time.

COST-EFFECTIVE SEARCHING ON ONESEARCH

After ten hours of online time, many of them using a demonstration password that did not display cost, I cannot give an experimentally proven answer. However, if you know what databases to use and are familiar with them and the nature of the search lends itself to a cross-file approach, then the most cost-effective way is to do your entire search in OneSearch. If you don't need to **SET DETAIL ON** or use **EXPAND**s or **SORT**s, OneSearch appears to be quicker, and therefore less expensive than single file searching. Cost differentials become more apparent as the search becomes longer. There was a 14% saving in a search in the business and engineering literature on the topic of specific expert system techniques applied to job shops or transportation.

Online cost is more difficult to estimate in OneSearch than in a single file search; it is more difficult to make adjustments if the search begins to get too expensive. If you do not know how many items you are getting from each file (you don't, unless **SET DETAIL ON** is in effect), you cannot estimate the cost of displaying the records. Time and cost saved online by avoiding **SAVE TEMP**s and **EXECUTE STEPS** is balanced by your preparation time. Time and cost saved by going offline to look for duplicates is balanced by the time you spend scanning the lists. There is a learning curve involved in using OneSearch, as with any new product. As you become more familiar with logical combinations and require less time to prepare, real savings may be realized.

However, in a final evaluation of OneSearch, cost is not the only issue.

CONCLUSION

Until a sort/merge option is added, the apparent advantage of using OneSearch is marginal. There are several tradeoffs to consider and compromises to make.

Ease of use online is counterbalanced by extra preparation time. Less time may be spent typing and re-executing the search online, but more time is required by the searcher offline before sitting down at the keyboard. The paper saved from not executing saved searches is used in **SET DETAIL ON**.

Precision is compromised by using free-text instead of controlled vocabulary and truncating terms so loosely as to ask for false drops.

Recall is compromised by using more general vocabulary rather than database-specific terms and related terms.

Flexibility is lost since it is difficult to adjust to the idiosyncrasies of different databases.

Before using OneSearch:

- Read your documentation
  - Check for different field labels
  - Check for different field entries
  - Check for different output formats
- Know your terminology
  - Check your terms for multiple meanings
  - Structure your search for alternative wording
- Evaluate your time vs. online time
  - Check costs carefully

What can DIALOG do to make OneSearch the searcher's real dream?

- Add a sort/merge function that flags duplicates
- Continue to reload databases with standardized field tags
- Encourage database producers to standardize field entry formats
- Have an online thesaurus like DBC's "Company Finder" to link terms
- Provide documentation, based on DIALINDEX categories, that identifies field tags, field entries, and output formats that are different within these categories
- Give users some free time to try OneSearch

While I cannot agree with the *CHRONOLOG*'s pronouncement that "OneSearch does it all!", I think DIALOG deserves credit for introducing OneSearch. Some of the problems with OneSearch can be solved by DIALOG. Some of the problems can be solved with cooperation from the database producers. And some of the problems require a willingness for searchers to compromise, be flexible, and try something new.

Communications to the author should be addressed to Ruth Pagell, Associate Librarian, Lippincott Library of the Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6207; 215/898-5922; DIALMIAL.

SEARCHING FULL-TEXT PERIODICALS: HOW FULL IS FULL?

by **Ruth Pagell**
Assistant Librarian
University of Pennsylvania

Editor's Note: This article and the one following it have opened a Pandora's box of issues surrounding full-text files and searching. The major online systems have added many full-text files, and refined their search software to enhance the searching of those files. Here is one expert searcher's attempt to understand the resulting chaos. We will welcome other readers' insights and contributions to the issues escaping that mythical box. —NG

Online full-text periodical files promise bibliographic instant gratification. Type in a few words or phrases and the complete texts of relevant magazine articles appear on your screen. Overnight document delivery, telefacsimile, and trying to find the article in a library cannot begin to compete with the promise of immediate availability of text in electronic form.

To learn more about full-text searching, I examined full-text periodical databases on four systems: BRS, DIALOG, Dow Jones, and Mead Data Central (MDC), with emphasis on Information Access Company's (IAC) ASAP files. *Financial World*, a periodical available on all four systems, was searched in depth. I wanted to answer several questions including:

- What type of information is available online in full text?
- How many periodical titles are available online in full text?
- How close is the match between the hardcopy versions of the periodical and the online records?
- How current are the online periodical files?
- What are the most efficient techniques for searching full-text periodical files on different systems?

TYPES OF FULL-TEXT FILES

Listed in Table 1 are some categories of full-text materials available on the systems I examined.

TABLE 1

TYPES OF FULL-TEXT FILES AVAILABLE

Category	Sample Titles
Business Magazines	*Forbes, Harvard Business Review*
Popular Magazines	*Time, Newsweek, Sports Illustrated*
Newspapers	*Wall Street Journal, New York Times*
Newswires	*AP, UP, PR Newswire*
Newsletters	*Chem Week Newswire, Platts Oilgram*

►

Reference Books	*Academic American Encyclopedia, Bible* (King James Version)
Legal Periodicals	*ABA Journal, Tax Lawyer*
Medical Journals	*Journal of American Medical Association*

NUMBER OF TITLES ONLINE

The number of full-text periodicals available online is growing, but is still small in comparison to the total number of journals indexed online. While MDC was once the prime source for full-text journals, database producers and publishers, like McGraw-Hill, IAC, Wiley, and now UMI/Data Courier, provide their own full-text files. However, there are still only a few hundred periodicals included in full-text files.

- Of about 1000 titles in **MAGAZINE INDEX** and **TRADE AND INDUSTRY INDEX**, only about 180 are currently in the ASAP files
- **McGraw-Hill Business Backgrounder** has 17 magazines
- UMI/Data Courier's **Business Dateline** has *selected* articles from 140 regional business publications
- MDC lists 56 individual magazines in its **MAGS** file. Many of these are on DIALOG in ASAP or Business Backgrounder.
- A DIALOG summer mailing listed about 355 "full-text" periodical titles. This overstates availability. Only selected articles from some journals are included: starting dates vary for individual titles: some of the listed titles are no longer online.

FULL TEXT ONLINE: IS IT REALLY COVER-TO-COVER?

Issues of 1985 *Financial World* were searched online and compared to the hardcopy versions (Table 2).

TABLE 2

1985 *FINANCIAL WORLD* RECORDS PER FILE

System	File	Number of Records
BRS	MSAP III	217
	TSAP III	177
DIALOG	647 (MSAP)	484
	648 (TSAP)	425
Dow Jones	Business Library	426
MDC	FINWLD	944

Coverage in MDC includes all articles, columns, special sections, and sub-sections. Individual tables are loaded as separate records. The number of records in FINWLD is much greater than in the Table of Contents of the hard copies. For example, for the March 20-April 2, 1985 *Financial World*, MDC has 48 individual records. Only 29 entries appear in the Table of Contents. IAC has 15 records for the same issue.

IAC files contain only the feature articles and columnists. The specially "boxed" sub-sections and tabular data have not been put into the file. In the March 20, 1985 issue, "Companies in the News," "People," and all the departments, are excluded from the IAC files.

Online titles vary from titles in the Tables of Contents. The March 20-April 2, 1985 issue of *Financial World* lists the following title:

> "Winning Converts: Investment advisers find some good values in convertible bonds"

The same article appears in MDC as:

> "Convertible Bonds: Watch Out for the Premium"

On IAC it is simply:

> "Watch out for the Premium."

Occasionally, an article appears as an online record that is not in the Table of Contents. Both online versions include a story entitled "The Saga of the Saatachi Boys," which does not appear in the journal's contents. This is a real problem only when using Dow Jones Business Library, since Dow Jones does not give page numbers.

ASAP III on BRS has only a portion of 1985 *Financial World*. The last issue of 1985 is missing from all IAC files.

Another peculiarity appeared in using *Financial World* as a search sample. Each issue is dated over a two-week time span, i.e., "Volume 154, Number 7; March 20-April 2, 1985." IAC enters that issue under the beginning date, March 20, and MDC under the final date, April 2.

CURRENCY ONLINE

One of the trade-offs between full-text and bibliographic files is currency. Exactly "how current" depends on the database and the system (Table 3).

Periodicals loaded as individual files on MDC generally have a one-issue lag over the newsstand copies (Table 4). Periodicals that are

TABLE 3

CURRENCY OF FINANCIAL WORLD and the ASAP FILES on 5/14/87

System	Database	Date File was Loaded	Most Recent *FW*	Most Recent ASAP Issue	# of 1987 Articles
BRS	MSAP III	2-87	1985	*Atlantic* 1-87	21
	TSAP III	3-87	1985	*Atlantic* 1-87	75
DIALOG	647(MSAP)	5-87	3-24-87	*PC Week* 4-14-87	5188
	648(TSAP)	3-87	1-6-87	*PR News* 2-26-87	7719
Dow Jones	Business Library	NA	1-20-87	*PR News* 3-30-87	10947
MDC	FINWLD	NA	6-2-87	*PC Mag* 4-14-87*	

* searched in MASAP

TABLE 4

MOST CURRENT ONLINE ISSUES COMPARED TO HARDCOPY*

Title	MDC	Hardcopy	Title	ASAP II on MDC	Hardcopy
Business Week	3-9-87	3-16-87	*Monthly Labor Review*	10-86	2-87
Economist	3-7-87	3-14-87	*Nation's Business*	12-86	3-87
Fortune	3-16-87	3-23-87	*Progressive Grocer*	9-86	3-87
Forbes	3-9-87	3-24-87	*Sales & Marketing Mgmt*	10-86	3-87

* Based on searches conducted 3/19/87 and availability in Lippincott Library on that date

TABLE 5

PROXIMITY SEARCHING AND FIELD DELIMITERS

Command	BRS	DIALOG	Dow Jones	MDC
Adjacency	**ADJ**[1]	**(W)**	implied[2]	**PRE/1**
Near	NO	**(2N)**	**NEAR2**	**W/2**
Same sentence	**WITH**	NO	**WITH**	NO
Same paragraph	**SAME**	**(S)**	**SAME**	NO
Same field	NO	**(F)**	NO	**W/SEG**
Specific field	Yes(**.TI.**)	Yes(**/TI**)	YES (**.HL.**)	**Section(ByLine)**
Expansion of term	**Root**	**Expand**	**Root**	No
Truncation	**$2**	**?? ?**	**$2**	**!,***[3]
Manipulate sets	Yes	Yes	Yes	No
Length of Article	Yes	Yes	No	Yes
Full-text output formats	Yes	Yes	Yes	Yes

[1] The **ADJ** operator is required in BRS native mode, but AfterDark and Menu have an implied adjacency

[2] An idiosyncrasy in searching Dow Jones is that adjacency is implied, unless there is another operator. For example **tax reform and legislation** retrieves **(tax reform) and legislation**, BUT **legislation and tax reform** retrieves **legislation and tax and reform**.

[3] Automatic singular/plural retrieval

►

part of the IAC databases can be as much as six months behind.

SEARCH STRATEGY

After searching the four systems, I concluded that differences in searching full text became a comparison of the software used by BRS, DIALOG, Dow Jones, and Mead Data Central. This confounded my attempt to isolate those issues unique to full-text online searching.

There are certain features in each system that are interesting to note when searching full text (Table 5).

Positive Features:

BRS: In displaying the output from a full-text search in BRS, the user can print an "occurrence table" which gives the location and frequency of the search terms within the full-text:

```
. .p bibl, oc/doc=1
```

Using `HITS`, those paragraphs and subparagraphs containing the search term can be printed:

```
. .p 1 hits/doc=1
```

* is a further refinement that allows the searcher to print only those sub-paragraphs that seem most relevant:

```
* ti, tx(5)/doc=1
```

DIALOG: Flexible searching and the `expand` command give the experienced searcher control over retrieval. KWIC output display, in files where it is available, is a useful feature.

```
?t3/3,kwic/1
```

Dow Jones: Dow Jones also has the capability to display only those text paragraphs where there is a match with at least one search term. `P*` is used after the first screen has been displayed.

`. .p/doc=1` prints the first screen.
`p*` then goes to the next screen with
a match.

MDC: Advantages in MDC include automatic singular/plural truncation, and KWIC display on all files. Also, files loaded by MDC include tables as well as text.

Major Deficiencies:

BRS: ASAP III is not as complete as documentation suggests. Not all titles for all dates are included. The "occurrence" table cannot be executed on the ASAP files.

DIALOG: KWIC is not yet available on many full-text files. As of the writing of this article, it is on MSAP but not TSAP.

Dow Jones: The number of periodical titles is limited. Article length and page numbers are not in the records.

MDC: There is a lack of flexibility in search strategy, combining sets, and manipulating terms. Controlled vocabulary is not an option.

CONCLUSIONS

I could spend years online and not analyze all of the permutations involved with full-text periodical searching. From the searching that I have done, the following conclusions stand out:

- Availability of full-text periodical titles is still a small percentage of all titles.
- ASAP files contain the largest number of full-text periodical titles, but variations in coverage and currency vary between MSAP and TSAP, and among the four systems.
- The definition of full text varies from selected stories, to major articles and columns, to cover-to-cover. Tabular and graphic data, as well as editorials, are often excluded.
- Periodicals loaded as separate files are often only one issue behind their print counterpart. Those that are part of larger files have a time lag of as much as six months.
- Powerful proximity operators and field delimiters, available in the command-driven systems, give the experienced searcher the flexibility to mix and match terms effectively.
- The major database systems are adapting their search software and output formats to enable better full-text searching.
- A comparison of the search results using full-text and controlled vocabulary files should be undertaken to dispel or substantiate the rumors about the quality of full-text output.

Are we getting full text online? What we are getting is the full text of articles from periodicals, but not an electronic duplicate of the magazine. Obviously, even those magazines that claim cover-to-cover online versions, cannot include graphics, photographs or advertisements. How full is full? The answer is dependent on the system and the database, but the glass is never filled to the brim.

Communications to the author should be addressed to Ruth Pagell, Assistant Librarian, Lippincott Library of the Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6207; 215/898-5922; DIALMAIL.

SEARCHING IAC'S FULL-TEXT FILES: IT'S AWFULLY CONFUSING

by **Ruth Pagell**
Assistant Librarian
University of Pennsylvania

IAC's **MAGAZINE INDEX** and **TRADE & INDUSTRY INDEX** and the companion **ASAP** files provide an innovative combination of bibliographic searching and full-text retrieval, or full-text searching and retrieval. The searcher can access MAGAZINE INDEX or TRADE & INDUSTRY INDEX, which contain citations from over 1000 magazines and newspapers. The full-text of selected articles from selected periodicals in these files are in **MSAP** and **TSAP**.

However, there are really seven different files on the four systems I examined—BRS, DIALOG, Dow Jones, and Mead Data Central (MDC) (Table 1).

TABLE 1

IAC FULL-TEXT FILES

SYSTEM	FILE	DATABASE LABEL
BRS	MAGAZINE ASAP III	MSAP
	TRADE & INDUSTRY ASAP III	TSAP
DIALOG	MAGAZINE ASAP	File 647
	TRADE & INDUSTRY ASAP	File 648
Dow Jones		Business Library
Mead (NEXIS)	ASAP Library Files	ASAP
	MAGAZINE ASAP II	MASAP
	TRADE & INDUSTRY ASAP II	TIASAP

The following discrepancies between expectations and performance are apparent:

1. The number of periodicals included in the files is different on all four systems. It is difficult to determine exactly how many periodicals there really are. IAC's printed list of publications, the list created from an online **expand** on DIALOG, and MDC's list reveal different titles (Table 2). Many periodicals have been dropped from the files. ►

TABLE 2
NUMBER OF TITLES IN IAC FILES

SYSTEM	MAGAZINE ASAP	TRADE & INDUSTRY ASAP
BRS	73	113[1]
DIALOG	110	209[2]
Dow Jones	NA	5[3]
MDC	80	119[4]

[1] Based on IAC documentation supplied in March, 1987

[2] Based on `EXPAND JN=` on files 647 and 648 on May 14, 1987

[3] From the current and previous banners on Business Library

[4] From NEXIS documentation dated October 1986.

2. File updates are loaded on different dates on each system.

3. Some periodicals appear in both MSAP and TSAP. However, the number of citations from these publications is different on the two files. I was surprised to learn that MSAP actually cited 174 *more* articles for the business magazine *Forbes* than TSAP.

Some periodicals appear in both MSAP and TSAP. However, the number of citations from these publications is different on the two files.

There are only 25 titles in MSAP on DIALOG that are not cited in TSAP. In order to reduce the guess work over whether to access MSAP or TSAP, MDC has created a "superfile" ASAP which merges the contents of the two files. IAC itself, or DIALOG and BRS, should consider doing this too.

4. Updates for the same title on the same system differ between the MSAP and TSAP files.

5. Titles and individual issues are missing from ASAP III, the BRS version.

6. On BRS, the full-text occurrence output format `OC` cannot be executed, because `OC` is a field label that generates an accession number.

7. Periodical titles may appear under variant spellings. For example, the following journal name entries appear in ASAP:

US News and World Report and *U.S. News and World Report*

SELECTION CRITERIA

There are two factors that determine inclusion of a periodical in an IAC full-text file.

1. Titles are selected based on potential usage. As a result of this decision, those periodicals likely to be in libraries, homes, or offices in print, such as *Forbes, Fortune*, or *Scientific American* are those available online.

2. Titles are included based on agreements with the publishers and the systems. McGraw-Hill has loaded **Business Backgrounder**, its own full-text database which includes *Business Week*, on DIALOG. None of the McGraw-Hill titles are available in the ASAP files.

Fourteen titles, including the Time-Life publications, are only on ASAP in DIALOG. *Forbes* is on this restricted list, even though the magazine has a non-exclusive license with IAC. Questioning initiated by this article revealed that IAC agreed to exclude *Forbes* from ASAP III on BRS at DIALOG's request!

ASAP on MDC excludes titles already available in NEXIS' own files. The MDC documentation clearly states this fact.

CONCLUSION

IAC ASAP files are useful for individuals looking for the full text of selected articles. IAC does not promise full coverage. However, exactly what articles from which periodicals you will retrieve depends on the version of ASAP that you use. These files cannot serve as substitutes for the journals themselves.

We deliver
the original article!

For everything abstracted or cited in CAS ONLINE, CA SEARCH, *Chemical Abstracts, CA Selects, Chemical Industry Notes* from the past 10 years (Soviet documents from the past 15 years) we'll deliver a copy or loan you the original:

- journal articles
- conference proceedings
- PATENTS (since 1978)
- technical reports

96% of document orders filled & shipped within 24 hours

Chemical Abstracts Service
Document Delivery Service
2540 Olentangy River Road
P.O. Box 3012
Columbus, OH 43210
800-848-6538

TABLE 3

JOURNALS IN IAC FULL-TEXT FILES

JOURNALS IN TSAP: DIALOG FILE 648 (May 16, 1987)

HITS	TITLE
627	A +
60	ADMINISTRATIVE MANAGEMENT (THE MAGAZINE OF)X
687	AIR TRANSPORT WORLD
14332	AMERICAN BANKER #
28	AMERICAN FEDERATIONIST
17173	AMERICAN METAL MARKET #
8	ATLANTIC M
479	AUTOMOTIVE INDUSTRIES
320	AUTOMOTIVE MARKETING
186	BACK STAGE
900	BACKSTAGE
185	BEST'S REVIEW – LIFE-HEALTH EDITION
336	BEST'S REVIEW – LIFE-HEALTH INSURANCE EDITION
506	BEST'S REVIEW – PROPERTY-CASUALTY INSURANCE ED
9	BOATING M*
476	BOATING INDUSTRY
1	BOY'S LIFE N*
2884	BROADCASTING
780	BUSINESS & COMMERCIAL AVIATION
1594	BUSINESS AMERICA
24	BUSINESS ATLANTIC
124	BUSINESS HISTORY REVIEW
134	BUSINESS JOURNAL-MILWAUKEE
159	BUSINESS JOURNAL-PORTLAND
57	BUSINESS JOURNAL-SAN JOSE
19	CALIFORNIA BUSINESS
1059	CANADIAN BUSINESS

JOURNALS IN MSAP: DIALOG FILE 647 (May 16, 1987)

HITS	TITLE
600	A + T
84	ADMINISTRATIVE MANAGEMENT (THE MAGAZINE OF) T
140	AGING *
91	AMERICAN EDUCATION
37	AMERICAN FEDERATIONIST
44	AMERICAN HERITAGE
706	AMERICAN LIBRARIES
1018	ATLANTIC
243	BACKPACKER *
635	BOATING *
85	BOYS' LIFE N*
1383	BUSINESS AMERICA T
1013	CANADIAN BUSINESS T
612	CAR AND DRIVER *
995	CHANGING TIMES
153	CHILDREN TODAY *
113	CO-ED *
437	COMPUTERS & ELECTRONICS T
786	CREATIVE COMPUTING *X
462	CYCLE *
1170	DATAMATION T
9006	DEPARTMENT OF STATE BULLETIN
238	DISCOVER X
1381	DUN'S BUSINESS MONTH T
1	DUN'S BUSINESS MONTHS
281	FDA CONSUMER
1675	FINANCIAL WORLD T

Key: #-Only on Dialog; X-not in IAC's Documentation; *-No longer included; @-Not on NEXIS IAC; M-Appearing in TSAP but on MSAP list; T in TSAP as well as MSAP; N- on NEXIS but not in IACs documentation

TABLE 3 (continued)

JOURNALS IN TSAP: DIALOG FILE 648 (May 16, 1987)

HITS	TITLE
748	CHAIN STORE AGE – EXECUTIVES EDITION
519	CHAIN STORE AGE – GENERAL MERCHANDISE EDITION
91	CHAIN STORE AGE – GENERAL MERCHANDISE TRENDS
251	CHAIN STORE AGE – GENERAL MERCHANDISING EDITIO
112	CHAIN STORE AGE SUPERMARKETS *
167	CHANGING TIMES M
81	CINCINNATI BUSINESS COURIER
677	COMMUNICATIONS NEWS
1191	COMPUTER DECISIONS
25	COMPUTER PICTURES X
247	COMPUTERS & ELECTRONICS
40	CONSTRUCTION REVIEW
609	CONSUMER ELECTRONICS
60	CRAIN'S CLEVELAND BUSINESS
1141	CREATIVE COMPUTING *
3	CYCLE M*
13491	DAILY NEWS RECORD #
1257	DATAMATION
500	DEFENSE ELECTRONICS @
2	DEPARTMENT OF STATE BULLETIN M
1472	DISCOUNT STORE NEWS
297	DISTRIBUTION
279	DISTRIBUTION WORLDWIDE X
369	DRUG & COSMETIC INDUSTRY *
963	DRUG TOPICS
1121	DUN'S BUSINESS MONTH @
1489	ELECTRONIC DESIGN
6	ELECTRONIC JOURNAL X
3410	ELECTRONIC NEWS
2792	ENERGY USER NEWS
214	EUROMONEY X
23	FDA CONSUMER M
21	FED. HOME LOAN BANK BOARD J. *
1649	FINANCIAL WORLD #
39	FLORIDA TREND MAGAZINE
1	FLOWER AND GARDEN *
4	FLYING *
404	FOLIO: THE MAGAZINE FOR MAGAZINE MANAGEMENT
1	FOOD & NUTRITION M

JOURNALS IN MSAP: DIALOG FILE 647 (May 16, 1987)

HITS	TITLE
246	FLOWER AND GARDEN *
641	FLYING *
74	FOOD & NUTRITION
6388	FORBES T
27	FORECAST FOR HOME ECONOMICS N
4773	FORTUNE T
599	GOLF MAGAZINE *
422	GUNS & AMMO *
696	HEALTH
1010	HIGH FIDELITY *
352	HOT ROD *
1188	INC. T
340	INDUSTRIAL RESEARCH T
2184	INDUSTRY WEEK T
102	JOURNAL OF SMALL BUSINESS MANAGEMENT TN
719	LADIES HOME JOURNAL *
136	MODERN BRIDE N
112	MODERN OFFICE PROCEDURES T
426	MODERN OFFICE TECHNOLOGY T
753	MODERN PHOTOGRAPHY *
2061	MONEY T
350	MONTHLY LABOR REVIEW T
247	MONTHLY REVIEW T
597	MOTOR TREND *
1813	NATION
1274	NATION'S BUSINESS T
490	NATURAL HISTORY *
98	NEA TODAY N*
266	NEW LEADER N
3158	NEW REPUBLIC N
84	OCCUPATIONAL OUTLOOK QUARTERLY T
683	OFFICE ADMINISTRATION AND AUTOMATION T
659	OUTDOOR LIFE *
2398	PC MAGAZINE X
4468	PC WEEK T
1	PC WEEK4
1	PC WEEK9
8624	PEOPLE WEEKLY
593	PETERSEN'S PHOTOGRAPHIC MAGAZINE *

2870	FOOTWEAR NEWS
248	FOOTWEAR NEWS MAGAZINE *
6214	FORBES #
4934	FORTUNE #
343	FOUNDRY MANAGEMENT & TECHNOLOGY
302	GIFT & DECORATIVE ACCESSORIES *
293	GIFTS & DECORATIVE ACCESSORIES *
188	GOVERNMENT PRODUCT NEWS
736	GRAPHIC ARTS MONTHLY
5	GUNS & AMMO M*
387	HANDLING & SHIPPING MANAGEMENT
416	HARDWARE AGE
32	HAWAII BUSINESS MAGAZINE
1961	HFD-THE WEEKLY HOME FURNISHINGS NEWSPAPER
3	HIGH FIDELITY M*
530	HIGHWAY & HEAVY CONSTRUCTION
522	HOME & AUTO *
319	HOSPITALS
1065	HOSPITALS, JOURNAL OF AMERICAN HOSPITAL ASSOCI
2	HOT ROD M*
534	HOUSEWARES *
46	HOUSTON BUSINESS JOURNAL
376	HYDRAULICS & PNEUMATICS
374	IMPLEMENT & TRACTOR
650	INC. @
340	INDUSTRIAL RESEARCH
2217	INDUSTRY WEEK @
7	INFORMATION PUBLISHING X
7	INFORMATION PUBLISHING: AN ELECTRONIC JOURNAL X
283	INFOSYSTEMS
559	INSTITUTIONAL DISTRIBUTION *
741	INTERIOR DESIGN
1197	IRON AGE
452	JEWELERS CIRCULAR KEYSTONE *
100	JOURNAL OF SMALL BUSINESS MANAGEMENT N
2	LADIES HOME JOURNAL M*
226	LAWN & GARDEN MARKETING *
1407	PLAYBOY *
946	POPULAR PHOTOGRAPHY *
1068	POPULAR SCIENCE
302	PSYCHOLOGY TODAY
876	RESEARCH & DEVELOPMENT T
1566	ROLLING STONE *
910	SALES AND MARKETING MANAGEMENT T
506	SATURDAY EVENING POST
27	SCHOLASTIC CHOICES N*
446	SCHOLASTIC UPDATE
5983	SCIENCE #
1629	SCIENCE NEWS #
156	SCIENCE'83
1	SCIENCE'83E
285	SCIENCE'84
1	SCIENCE'84D
1	SCIENCE'84H
1	SCIENCE'84O
1	SCIENCE'846
209	SCIENCE'85
1	SCIENCE'85E
34	SCIENCE'86
1099	SCIENTIFIC AMERICAN
10	SCOUTING N*
476	SKIING *
752	SKIN DIVER *
548	SMITHSONIAN
67	SPORT X
4234	SPORTS ILLUSTRATED #
2652	STEREO REVIEW
3033	SUNSET
614	TECHNOLOGY REVIEW
479	TEEN *
11094	TIME #
6000	U.S. NEWS & WORLD REPORT T
750	UN CHRONICLE
572	UNESCO COURIER

Key: #-Only on Dialog; X-not in IAC's Documentation; *-No longer included; @-Not on NEXIS IAC; M-Appearing in TSAP but on MSAP list; T in TSAP as well as MSAP; N- on NEXIS but not in IACs documentation

TABLE 3 (continued)

JOURNALS IN TSAP: DIALOG FILE 648 (May 16, 1987)

HITS	TITLE
7	LIFE M
293	LODGING HOSPITALITY *
5	LOS ANGELES BUSINESS JOURNAL
1202	MACHINE DESIGN
868	MARKETING & MEDIA DECISIONS
193	MEDICAL ECONOMICS X
469	MEDICAL LABORATORY OBSERVER
416	MEETINGS & CONVENTIONS *
198	MERCHANDISING
12	MINNESOTA BUSINESS JOURNAL
111	MODERN OFFICE PROCEDURES X
421	MODERN OFFICE TECHNOLOGY
3	MODERN PHOTOGRAPHY M*
408	MODERN TIRE DEALER *
2113	MONEY *
330	MONTHLY LABOR REVIEW
225	MONTHLY REVIEW N
168	MOTOR AGE *
9	MOTOR TREND M*
75*	NATION M
1230	NATION'S BUSINESS
2718	NATION'S RESTAURANT NEWS
743	NATIONAL PETROLEUM NEWS
15	NATIONAL REVIEW M
98	NEA TODAY N*
58	NEW ENGLAND BUSINESS
20	NEW LEADER N
266	NEW REPUBLIC NM
167	NURSING HOMES N
64	OCCUPATIONAL OUTLOOK
16	OCCUPATIONAL OUTLOOK QUARTERLY
707	OFFICE ADMINISTRATION AND AUTOMATION
6966	OIL DAILY #
1	OUTDOOR LIFE M*
478	PATIENT CARE
2003	PC
208	PC MAGAZINE X
4163	PC WEEK

JOURNALS IN MSAP: DIALOG FILE 647 (May 16, 1987)

HITS	TITLE
95	WASHINGTON MONTHLY
157	WHOLE EARTH REVIEW
270	WORKBASKET
455	WORKBENCH *
1200	WORKING WOMAN T
296	WORLD HEALTH *
15	YOUR PUBLIC LANDS
25	current, primary, unique titles

50	PEOPLE M #
8	PEOPLE WEEKLY X
2	PETERSEN'S PHOTOGRAPHIC MAGAZINE M*
116	PETS-SUPPLIES-MARKETING *
146	PHILADELPHIA BUSINESS JOURNAL
79	PITTSBURGH BUSINESS TIMES-JOURNAL
618	PLANT ENGINEERING
9	PLAYBOY M*
421	PLAYTHINGS
1	POPULAR PHOTOGRAPHY M*
12	POPULAR SCIENCE M
168722	PR NEWSWIRE #
383	PRODUCTION ENGINEERING
683	PROGRESSIVE ARCHITECTURE
710	PROGRESSIVE GROCER
14	PSYCHOLOGY TODAY M
272	QUICK FROZEN FOODS *
897	RESEARCH & DEVELOPMENT
1089	RESTAURANT BUSINESS MAGAZINE
704	RETAILING HOME FURNISHINGS X
544	RN
11	ROLLING STONE M*
51	SALES & MARKETING MANAGEMENT
928	SALES AND MARKETING MANAGEMENT
25	SAN DIEGO BUSINESS JOURNAL
68	SAN FRANCISCO BUSINESS JOURNAL
7	SATURDAY EVENING POST M
18	SCHOLASTIC UPDATE M
324	SCHOOL PRODUCT NEWS *
41	SCIENCE M#
42	SCIENCE NEWS M#
2	SCIENCE'83 M
1	SCIENCE'84 M
2	SCIENCE'85 M
1	SCIENCE'86 M
17	SCIENTIFIC AMERICAN M
3	SKIING M*
83	SKIING TRADE MONTHLY NEWS

Key: #-Only on Dialog; X-not in IAC's Documentation; *-No longer included; @-Not on NEXIS IAC; M-Appearing in TSAP but on MSAP list; T in TSAP as well as MSAP; N- on NEXIS but not in IACs documentation

TABLE 3 (continued)

JOURNALS IN TSAP: DIALOG FILE 648 (May 16, 1987)

HITS	TITLE
332	SKIING TRADE NEWS *
4	SKIN DIVER M*
5	SMITHSONIAN M
334	SOAP-COSMETICS-CHEMICAL SPECIALTIES
77	SOC. SEC. BULL.
17	SOCIAL SECURITY BULLETIN
2	SPORT X
22	SPORTS ILLUSTRATED M#
40	ST. LOUIS BUSINESS JOURNAL
35	STEREO REVIEW M
494	SUPERMARKET BUSINESS N
3095	SUPERMARKET NEWS
184	SURVEY OF CURRENT BUSINESS
532	TECHNOLOGY REVIEW
1308	TELEPHONE ENGINEER & MANAGEMENT
1374	TIME M#
366	TOOLING & PRODUCTION *
6503	TRAVEL WEEKLY
139	U.S. NEWS & WORLD REPORT @
4	UN CHRONICLE M
1	UNESCO COURIER M
1288	US NEWS AND WORLD REPORT
574	WARD'S AUTO WORLD
200	WASHINGTON BUSINESS JOURNAL
10	WASHINGTON MONTHLY NM
3	WHOLE EARTH REVIEW M
11247	WOMENS WEAR DAILY
904	WORKING WOMAN
1	WORLD HEALTH M*
12	YACHTING M*

121 current,primary,unique titles

Key: #-Only on Dialog; X-not in IAC's Documentation; *-No longer included; @-Not on NEXIS IAC; M-Appearing in TSAP but on MSAP list; T in TSAP as well as MSAP; N- on NEXIS but not in IACs documentation

Communications to the author should be addressed to Ruth Pagell, Assistant Librarian, Lippincott Library of the Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6207; 215/898-5922; DIALMAIL.

HOW TO USE INFORMATION SEARCH TACTICS ONLINE

by Marcia J. Bates
Associate Professor
Graduate School of Library and Information Science, UCLA

INTRODUCTION

In earlier articles I have described search tactics, i.e., moves made to further a search, that can be used in all kinds of information searching, both manual and online [1,2]. Some people have told me that they found the tactics intriguing but wanted more guidance in their use online. So for this article I have selected tactics which are particularly useful for online searching, will discuss how to use these tactics, and will identify the points in a search at which they are most useful. At the end of the article an extended example will show how all the tactics may be used in an online search.

Definitions. A *search tactic* is a move made to further a search. Such moves may be good, i.e., promote search effectiveness, or bad. It is hoped that the tactics recommended here are examples of good moves. The searcher must still exercise judgment, however, in deciding when to use a tactic; even good tactics can be ineffective if used inappropriately.

A *search strategy* is a plan or approach to an entire search. The term search strategy is often confused with *search formulation*, which is the statement or set of statements which express the query in a form understandable to the online service. A search formulation is simply an expression (among many possible) of the search query, whereas a search strategy is an overarching plan and operates above specific term choice and command use. A particular search formulation may arise out of a search strategy that the searcher has in mind, or may simply be just a way of getting the query into the system, without much of a plan guiding the input. An example strategy is Charles Bourne's "Building Block" strategy [3, p. 12-13], in which the searcher arranges the topic of the search into ORed sets, inputs the sets separately, then later combines them with ANDs.

TACTICS TO USE EARLY IN THE SEARCH

BIBBLE. To BIBBLE (named after "bibliography") is to see if the searching one plans to do has already been done by someone else, especially the preparation of a bibliography or

►

a literature search. Although one's first impulse may be to launch into an extensive search when a client asks for information, such an effort is unnecessary if the work, or part of it, has already been done. The essence of BIBBLE is to let your predecessor do the work.

Many of the articles, reports, or books covered in online databases are themselves extensive bibliographies or state-of-the-art reviews. In the sciences, a "review" often refers to a survey of a large body of current research literature, rather than to the evaluative reviews of a single book found in literary sources. The presence of "review" in the title of a scientific journal often indicates that the journal concentrates on publishing reviews of this type.

Databases vary in how well they identify bibliographic sources. Some routinely assign terms like "review" or "bibliography" as part of the indexing. In such cases, one need only identify these bibliographic terms, input them all as an **OR**ed set (known as a "hedge"), and then **AND** them with the topic of interest to get a listing of all the bibliographic materials in the database on one's topic. For example, one might create a saved search on an **OR**ed set of terms such as: **Bibliography OR Review OR State of the art**, plus any other bibliographically rich terms in that database. (Sometimes these bibliographic terms appear as document type codes instead of descriptors.) Then, the first thing one does at the beginning of all searches involving any kind of survey of the literature is to call up that BIBBLE search and **AND** it to the terms representing the topic of the current search. Thus, one can very quickly run a search which produces *other people's* lengthy searches, i.e., their literature reviews or bibliographies.

If the database is not indexed by terms indicating bibliographically rich items, then it will be necessary to look for these materials in other ways. For example, one could **AND** the topic terms with the title of one or more journals that routinely carry review articles.

Once one or more bibliographies or reviews on the topic are found, they may be retrieved from the local library or information center and, depending on their coverage, they may constitute the bulk of the material needed to meet the requester's need. Then the rest of the search need only involve updating the found bibliography or review up to the present, and/or doing some supplemental searching if the topic of the found item(s) does not match exactly with what one's client wants—a much cheaper proposition than conducting a full search and having it printed out.

In an organization that does a lot of online searching, records kept of searches and who received them will make local "bibbling" possible. Either the actual printout of earlier searches may be located or previously downloaded records may be retrieved. (Downloading must be done with appropriate legal arrangements with online services and database producers.)

The tactic **RECORD** is to keep track of trails one has followed. An early study of librarians' search strategy found that unless they made a note to themselves, they did not follow up discrepancies or other problems they encountered in searching[4].

A somewhat different aspect of RECORD appears in the following experience told to me by a searcher in a large industrial organization. In doing a search for a researcher in the organization, she had great difficulty formulating the search, revised it several times and still was not happy with it. Finally, she felt that she had done the best she could, sent on the results, and tried to forget all about it. A few days later she got a note from him; he was ecstatic with her search and wanted her to perform update searches for him on the topic regularly. The only trouble was, she had not kept a record of her search formulation.

SURVEY: This tactic is to review, at each decision point of the search, the available options before selecting. For example, before beginning an online search, one's natural inclination may be to start searching on the "obvious" database, and to use only the terms supplied by the requester. On the other hand, it might take only a few seconds to review what other databases are available that are relevant to the topic. In surveying these other options, one may find, for example, that a foreign, full text, or citation database would be more suitable for the designated query than the ones usually consulted. The little bit of effort spent reviewing these other possibilities can eliminate expensive search time spent on a database that has some but not the most relevant materials.

SURVEY is also a valuable technique for selecting search terms. At first only one or two terms may seem to apply to your topic, tempting you to jump right in with them only. But a moment's reflection, or a look at the cross-references in a thesaurus, will often produce several more terms of value. If one uses SURVEY, the search formulation can more effectively express the search topic *from the start*, and valuable time can be saved.

One of the main tasks early in an online search is to identify good terms to search with. If one searches on only one or two apparently relevant terms, and gets back a postings set of the "right" size, it is tempting to stop right there and give the requester the output set as is. (See reference[5].) But such searches are often inadequate because they miss the best terms for the topic. So, except where the search is specifically limited for some reason, it is best to manipulate the search formulation until it appears to be the best for the stated need.

Several tactics can be used to generate additional terms to enrich the search formulation. To **NEIGHBOR** is to find additional search terms by looking at neighboring terms in alphabetically arranged lists – a familiar technique in online information systems using commands such as **Expand** (DIALOG) and **Root** (BRS). To **TRACE** is to examine information already found in the search in order to find additional terms to be used in furthering the search. In manual searching one looks at the tracings on a catalog card; in online searching one examines the terms in the descriptor lists and abstracts of sample citations printed out.

Three additional tactics for generating search terms involve moving about in the hierarchy, or "family tree" of subject terms between broader (more general) terms and narrower (more specific) ones. It is important to remember that a given topic 1) may be discussed in documents wholly devoted to that topic, 2) may appear as a subtopic in a document devoted to a broader topic, or 3) may have just part of its content dealt with in a document on a narrower topic. Documents of all of these three types may be of interest to the requester, but documents of only the first type will be retrieved if the only search terms used are those that are exactly descriptive of the topic, and not broader or narrower. Thus, it is a good idea to try to locate terms to search with that are on several levels of specificity relative to the topic of interest.

One can expand or alter a search formulation by using three tactics for moving around the hierarchy of broader and narrower terms:

SUPER:	To move upward hierarchically to a broader (superordinate) term.
SUB:	To move downward hierarchically to a more specific (subordinate) term.
RELATE:	to move sideways hierarchically to a coordinate term.

The searcher can perform these tactics 1) mentally, by thinking up broader, narrower, etc. terms, 2) by searching through a manual thesaurus that lists BT (broader term), NT (narrower term), and RT (related term), (or similar terminology), or 3) by following up related terms that are provided in some online database indexes.

An example of the third case can be found with **ERIC** on DIALOG. In ERIC, using the **EXPAND** command in different ways can either produce the terms that are alphabetically close to the entry term, or the terms that are *conceptually* related to the entry term. In the latter case, a code indicates whether the term is broader, narrower, or related (coordinate) to the entry term.

Two other tactics are helpful as well in identifying search terms. Every searcher is familiar with the case where a widely used term mysteriously produces no postings, or only a very small number – far fewer than could reasonably be expected for that term in that database. Usually this occurs because the database uses a spacing or spelling variant other than the one that occurred to the searcher to use. "Poly vinyl chloride," for example, may appear as three words, two words, one word, or as "PVC." Sometimes the situation is the reverse, and the version the searcher has used is the most popular one – but a few other documents may also be found under less common spelling or spacing variants.

The tactics **RESPELL** and **RESPACE** are, respectively, to look for spelling and spacing variants of the terms used in one's search formulation. These tactics are good to keep in mind at all times, and especially when there are suspiciously few postings under a given term.

TACTICS TO USE LATER IN THE SEARCH

Combining search terms with Boolean logic is at the heart of online searching. In order to search effectively using Boolean logic, it is important to be aware of the difference between a *term* and a *concept*, since the two are treated differently in search formulations. A concept may be expressed by several synonymous or nearly synonymous terms, each of which, in turn, may contain one or several words. For example, the single *concept* of a type of educational resource may be expressed by two different two-word *terms*, "Instructional media" or "Audiovisual aid."

In online searching, *concepts* are generally combined using the Boolean operator **AND**, and *term variants* are combined with the operator **OR**. For a search on the topic of "Creationism in high school biology curricula," four concepts can be identified: creationism, biology, curriculum, high schools, which we would expect to **AND** together.

One might wish to express one or more of these concepts with more than one term. For example, creationism may show up as a topic of discussion in documents on science curricula in general (and include mention of biology) or in discussions of zoology curricula (the part of biology most likely to be concerned with creationism). So the biology component could be expanded with additional terms to read: **Biology OR Science OR Zoology**. One might think up these alternate terms by employing

►

NEIGHBOR, SUPER, and the other term-generating tactics described above.

With this basic pattern in mind of ORing term variants of a concept and ANDing concepts together, one can see that several search tactics can be used to manipulate the search formulation until it produces the sort of postings set desired. First, here are the tactics:

- To **EXHAUST** is to include most or all of the concepts of a search topic in the initial search formulation, or to add one or more of the query concepts to an already-used search formulation.
- To **REDUCE** is to minimize the number of concepts of a search topic in the initial search formulation, or to subtract one or more of the query concepts from an already-used search formulation.
- To **PARALLEL** is to make the search formulation broad (or broader) by including, or adding, synonyms or otherwise conceptually parallel terms.
- To **PINPOINT** is to make the search formulation precise by minimizing, or reducing, the number of parallel terms, retaining the more perfectly descriptive terms.
- To **BLOCK** is to reject items indexed by certain terms. In online searching, BLOCK is equivalent to **AND NOT**.
- To **FIX** is to try alternate affixes, i.e., alternate prefixes, infixes, or suffixes. This tactic is implemented in online search systems as truncation.

Through use of these tactics, the searcher can manipulate the search formulation in powerful ways—to increase or decrease the output, and to improve precision or recall. Specifically, one can add in additional conceptual elements by ANDing them into the search formulation (EXHAUST), or one can REDUCE the number of concepts by dropping ANDed elements. In the creationism example cited above, one could REDUCE the query by subtracting any one of the four concepts, or, if one had earlier input only two or three of the concepts, one could add concepts (EXHAUST).

> **Through use of these tactics, the searcher can manipulate the search formulation in powerful ways—to increase or decrease the output, and to improve precision or recall.**

To BLOCK is to NOT a term out of a search formulation. NOT is really short for AND NOT, i.e., NOT is another form of AND. Consequently BLOCK has a similar impact to that of EXHAUST; to add a NOT to a search formulation is like adding another concept. In searching a NOTed term, the system first finds all documents in which the term appears (just as it would for an AND), but then drops all documents containing the term rather than including them as it would for an AND.

Next, any given conceptual element can be broadened (a wider net cast) by ORing in additional variant terms (PARALLEL) into one or more of the concepts. Alternatively, the search formulation can be made more precise and specific by dropping one or more term variants from any given concept (PINPOINT). In the example above, if one started only with "Biology" for one of the concepts, to PARALLEL would be to OR in "Science" or "Zoology" or both, with PINPOINT, of course, being to remove one or more of these terms from that conceptual element of the query.

Another way to create additional variant terms is through use of FIX, or truncation. To search on the truncated term "Librar?" is actually implicitly to search on an ORed set of all the terms starting with that root, for example, **Library OR Libraries OR Librarianship**, etc. So FIX works in a manner similar to that of PARALLEL—it tends to broaden the search.

The more concepts one ANDs in, the more stringent the search formulation will be, i.e., the more concepts one requires be present (EXHAUST) or absent (BLOCK), the fewer documents are likely to be found that contain, or exclude, all those terms. Similarly, the fewer variant (ORed) terms one lists for each concept (PINPOINT), the fewer documents there will likely be that match the request. So EXHAUST, BLOCK, and PINPOINT, each working in a different way, tend to *decrease* the number of postings produced by a search formulation. Such precise and specific search formulations are likely to increase precision and reduce recall.

On the other hand, REDUCE, PARALLEL, and FIX, each in a different way, tend to *increase* the number of postings returned from a search formulation. Fewer concepts (REDUCE) and more term variants for each concept (PARALLEL, FIX) tend to improve recall and reduce precision.

If one finds, late in the search, that too many or too few postings are coming back, tactics mentioned under "Early in the Search" can also be used at this stage. Narrower, more specific terms tend to have fewer postings, because terms of narrow scope are usually applicable to fewer documents. So if one is retrieving too many postings, SUB is a good tactic to use. That is, either with the help of system cross-references or on one's own, one should try to identify narrower terms to use than the ones already appearing in the search formulation.

On the other hand, if too few postings are appearing, RELATE and SUPER may help one find additional terms or broader terms to use. Also, any time a term draws very few documents, especially under three, then one should suspect that variant spelling or spacing is preferred by the database and one should try RESPELL or RESPACE.

Finally, as noted earlier, these tactics were originally developed to be techniques that the searcher could use regardless of whether the search was being conducted in a manual or online system. Thus, two specifically online-based tactics were not discussed: use of field delimiters and use of free-text. In general, limiting the search to matches on specific fields improves precision, while the use of free-text may improve both precision and recall, depending on the circumstances. Since this article is on the above-mentioned tactics, these techniques of free-text searching and field delimiting will not be discussed further, but should be kept in mind in practice for searching.

So, in sum, the following tactics are recommended for the following situations:

- Too many postings, try:
  EXHAUST, PINPOINT, BLOCK. Find more specific terms to search on by using SUB. Also, if truncation has already been used (FIX), remove it.
- Too few postings, try:
  REDUCE, PARALLEL, FIX. Generate more terms for PARALLEL by using NEIGHBOR, TRACE, RELATE, SUPER.
- No postings, try:
  RESPACE, RESPELL, plus the tactics listed under "Too few documents."

A SEARCH EXAMPLE USING THE TACTICS

Recall that a tactic is a move made to further a search; a tactic, or even a cluster of tactics, is not an overall strategy for a whole search. So in the example search to be described below, while tactics are demonstrated at the point in the search where they are likely to be useful, *the particular sequence or combination of tactics below should not be taken as a recommended overall strategy.* In practice the searcher picks and chooses which tactics to use when and where, depending on the particular query and needs at the time. In fact, I have deliberately used all the tactics I could in order to demonstrate their use; in practice you are more likely to use just some of them in any given search.

One final point about the example that follows. Many different factors can influence recall and precision in a search, and one can vary a search in several ways at once, with possibly mixed impact on the outcome. In order to keep down confusion and to demonstrate the impact of particular tactics, the search below will use only the tactics being demonstrated at each stage. For simplicity's sake, search terms will be limited to descriptors. Use of free-text, field delimiters, and "LIMIT" commands, while very common in practice, will be left out in order to demonstrate one effect at a time. Keep in mind, however, that the tactics can be used with these other capabilities as well.

Let us turn now to the example search:

OFFLINE. Let us suppose that we have the following search query: "I am interested in the social adjustment problems of the deaf and partially hearing when they are mainstreamed in high school." Offline, we would first analyze the query into its distinct concepts:

Concept 1	**Concept 2**	**Concept 3**	**Concept 4**
Deaf	Mainstreaming	Social adjustment	High school
Partial hearing			

As noted earlier, we formulate searches by **AND**ing distinct concepts and **OR**ing term variants. So, taking what we have so far, our search formulation would look like this:

(Deaf OR Partially hearing) AND Mainstreaming AND Social Adjustment AND High school

To be deaf and to be partially hearing are, of course, different conditions; these terms are not synonyms. But *for our purposes in this search*, they can be seen as variations on the same concept, because we are interested in documents discussing either or both conditions.

Now that we have a sense of the query, we want first to SURVEY which database to search. **ERIC** might be the first one to come to mind here, but a moment's thought – perhaps helped by examining an online service's database catalog – may suggest several other good possibilities. On reflection, we realize that other possibilities are **Exceptional Child Education Resources**, **PsycINFO**, **MEDLINE**, **LC MARC**, **Books in Print**, **Social SciSearch** (either the citation or "title term" portions of the database), and possibly others. In practice, depending on the needs of the client, we might choose one or several of these other databases.

►

Here, because it is so well known, we will use ERIC in the example.

Next we SURVEY terms to use in the search formulation. Looking in the *Thesaurus of ERIC Descriptors* we find that *Deafness, Partial hearing*, and *Hearing Impairments* are the descriptors to use both for these conditions and for the people having them. We find also that *Mainstreaming, Social adjustment*, and *High schools* are the valid terms for these concepts. We also note that *Secondary education* deals with both junior high schools and high schools and might also cover material of interest. So now our modified search formulation looks like this:

```
(Deafness OR Partial hearing OR Hearing Impairments)
AND Mainstreaming AND Social adjustment AND (High schools OR
Secondary education)
```

ONLINE. We have learned from experience, however, that no matter how carefully we review possible descriptors in thesauri, there are often other descriptors that did not occur to us to look up. The best way to discover these other descriptors is to use the TRACE tactic. We can do this by doing what has been called a "Brief-search"[3, p. 8] online. We input just the essential terms in the query, without worrying about formulating a complete, ideal set of search statements, and then have a few citations printed out. In this case suppose our brief search formulation is the following:

```
Mainstreaming AND Hearing Impairments
```

We then use format 8 in DIALOG to print out just titles and indexing of a half dozen records. At this point we just want to find more search terms so this format best meets that need. Examining the descriptor lists of these records we identify three more useful terms for the "Social adjustment" concept: *Interpersonal competence, Emotional adjustment*, and *Peer acceptance*.

Suppose that *Emotional adjustment* is a term that we had not thought to look up in the thesaurus while offline. So now, online, we decide to see if there are more useful terms related to this term. On DIALOG's ERIC we can NEIGHBOR this term by inputting:

```
Expand Emotional adjustment
```

We will then get an alphabetical listing of descriptors and free-text terms that begin with the word "Emotional":

```
Ref    Items    RT    Index-term

E1     11403           Emotional
E2         2           Emotional abuse
E3      1265    11    •Emotional adjustment
E4         0     1     Emotional behavior
E5         1           Emotional concepts
E6         1           Emotional content
E7         1           Emotional control card
E8      1165    15     Emotional development
E9         3           Emotional distress
E10     3353    26     Emotional disturbances (persistent,
                      serious emotional disorders and. . .)
E11      456     6     Emotional experience
E12        1           Emotional expression
```

Alternatively, we can input the same statement enclosed in parentheses:

```
Expand (Emotional adjustment)
```

and receive back a listing of the *conceptually* related terms, i.e., be shown a portion of the *Thesaurus of ERIC Descriptors*:

```
Ref    Items    Type   RT    Index-term

R1      1265           11    •Emotional adjustment
R2         0     U      1    Emotional maladjustment
                             (1966 1980)
R3      4129     B     33    Adjustment (to environment)
R4       109     R     10    Adaptive behavior (of
                             disabled)
R5       550     R     14    Affective measures
R6      1742     R     19    Alienation
R7      3353     R     26    Emotional disturbances
R8       900     R     17    Emotional problems
R9      1695     R     15    Morale
R10      797     R     15    Psychiatry
R11      889     R     18    Psychopathology
R12     1941     R     34    Psychotherapy
```

This capability in ERIC enables us to exercise the term generation tactics of SUPER, SUB, and RELATE. The third column from the left above states whether a term is broader ("B"), narrower ("N"—not in this particular listing), or related ("R") to our initial term. So in one listing we can go in all three directions to find additional, possibly relevant terms for the query.* In this case, we identify two new terms for the social adjustment concept of the query: *Adjustment(to environment)* and *Adaptive behavior(of disabled)*. Adding in these new terms, the query now has the following look:

```
(Deafness OR Partial hearing OR Hearing Impairments) AND
        Mainstreaming AND (Social adjustment OR Interpersonal com-
        petence OR Emotional adjustment OR Peer acceptance OR Ad-
        justment to environment OR Adaptive behavior (of disabled))
        AND (High schools OR Secondary education)
```

Inputting this query we get six documents back. We have a few printed out and they seem promising. It is possible, however, that some other combination would be still more promising, and we shall check out various other combinations of these elements.

Before doing so, however, we want to see if there are helpful literature reviews or state of the art papers that would cover all the needed literature until recently, in other words, we want to use the BIBBLE tactic. If we have searched regularly on ERIC, then we would have created a "hedge" or `OR`ed set of terms relating to bibliographies and literature reviews which we can now call up as a saved search. If not, then we can create the hedge now.

Let us suppose we have two BIBBLE hedges in ERIC, 1) a long high-recall one covering bibliographies and literature reviews of every type in both descriptors and document codes, such as the following:

---

* The "U" in front of "Emotional maladjustment" stands for "use for," i.e., use the entered term, "Emotional adjustment" instead of "Emotional maladjustment." Note that the latter term is therefore not legitimate; it has no postings.

DT=070 OR DT=071 OR DT=131 OR Annotated bibliographies OR Citations (references) OR Literature guides OR Literature reviews OR Research reviews OR State of the art reviews

and 2) a short, high-precision one such as the following:

Literature reviews OR State of the art reviews

At this point it is not clear which BIBBLE hedge will work out best nor how detailed we should be in the description of the topic. For example, while we are interested in the specific topic of the social adjustment problems of the mainstreamed deaf and partially hearing in high school, it is quite possible that valuable reviews would exist at other levels of specificity. For example, a book-length review covering all of mainstreaming could have a lengthy subsection on mainstreaming of hearing impaired, and thus be quite valuable even though its topic is far broader than our interest.

So we experiment with various combinations of terms in the subject query and with the two BIBBLE hedges until the printout of the first half dozen or so records looks highly promising. The final formulation we settle on is the following:

Mainstreaming AND (Deafness OR Hearing impairments OR Partial hearing) AND (Literature reviews OR State of the art reviews)

This BIBBLE search produces several promising records including the following:

"Exceptional Pupils: A Review of the Literature" (1981, 338p.)

"Research and Development Concerning Integration of Handicapped Pupils into the Ordinary School System" (1980, 176p.)

With any luck, when we examine them, these two lengthy reviews will suffice to cover the literature up to about 1980 so that subsequent searching can be limited to post-1980 literature.

Now let us return to the search proper. We have used various tactics to identify good search terms as well as earlier literature reviews so we do not have to repeat other people's work. We have thus applied most of the tactics appropriate to the early part of the search. Now let us experiment with tactics described for later in the search.

Earlier, it was said that EXHAUST and REDUCE are opposites, with the former tending to promote precision and the latter recall. These tactics involve increasing and reducing, respectively, the number of **AND**ed concepts in a search formulation. With our example search, then, let us see what the impact is of changing the number of **AND**ed concepts. Rather than repeating all the term variants below, let us simply call the concepts the following names: Mainstreaming, Hearing disability, Adjustment, and Education. *(Keep in mind that each component contains the full number of term variants appearing in the search formulation above, so, for example, the Adjustment component contains six term variants, "Social adjustment," "Interpersonal competence," etc.)* The following concept combinations were actually searched on DIALOG ERIC (like all the other examples in this section) and produced the number of postings given below.

Mainstreaming	**4107**
Mainstreaming AND Hearing disability	**328**
Mainstreaming AND Hearing disability AND Adjustment	**41**
Mainstreaming AND Hearing disability AND Adjustment AND Education	**6**

It can be seen that whichever one of these Boolean combinations we had started with, if we had added **AND**ed concepts (EXHAUST) the size of the output would go down and if we had dropped **AND**ed concepts (REDUCE), the size of the output would rise. In fact, adding **AND**s always leads to sets that are the same size or smaller, and dropping **AND**s always leads to sets that are the same size or larger. This fact holds true in the nature of Boolean logic, not just in practical experience. Similarly, since, as noted earlier, BLOCK is another form of **AND**, each **NOT**ed term added leads to sets of the same or smaller size.

We can see the effects of PARALLEL and PINPOINT in a similar way:

Emotional adjustment	**1265**
Emotional adjustment OR Social adjustment	**2432**
Emotional adjustment OR Social adjustment OR Adaptive behavior (of disabled)	**2531**

It can be seen here that when we add **OR**ed term variants (PARALLEL), the size of the output will stay the same or be larger and when we drop **OR**ed term variants (PINPOINT) the output will stay the same or be smaller. This fact, too, holds true in the mathematical nature of Boolean logic, not just in practice. Similarly, to use truncation (FIX) in effect **OR**s in additional terms, so FIX and PARALLEL have a similar effect. For example:

Deafness	**2444**
Deaf?	**3756**

Now that we have seen the effects of each of these tactics one at a time, let us return to a more realistic sort of search modification and see how the tactics might work in practice with the example search. Suppose that, fearing a high output, we had begun the search with a high precision formulation, that is, we had included all the distinct concepts we wanted and used only a few term variants:

Mainstreaming AND (Deafness OR Partial hearing) AND Social adjustment AND High schools NOT Junior high schools

▶

With this formulation we ran the risk that in **NOT**ing junior high schools, we might eliminate some documents that discussed both high schools and junior high schools. But if, fearing very large sets, we eliminated mixed documents, i.e., ones discussing both topics, we would be sure to get only documents wholly devoted to the topic, a reasonable choice if high recall was not needed. The above search, run on DIALOG ERIC, produced a null set, however. So let us use REDUCE and drop the **NOT** Junior high schools.

Mainstreaming AND (Deafness OR Partial hearing) AND Social adjustment AND High Schools

We still get a null set. Now we can REDUCE the formulation by dropping yet another element, and/or try using PARALLEL and adding in additional **OR**ed terms in the remaining elements. Using the *Thesaurus of ERIC Descriptors* on or offline, we might at this point discover the term variants for the deafness and social adjustment aspects, as well as the term *Secondary education*. **OR**ing those into the formulation (PARALLEL) would get us the six documents found earlier for the full search formulation. Seeking to increase this number we might drop the education element (REDUCE), leaving only the mainstreaming, disability, and adjustment concepts, on the assumption that any discussion of mainstreaming might still be of value, whether or not in a school context. Doing so, we would get back the 41 postings found above in the example where we **AND**ed one, two, three, and four elements. As noted earlier, still other capabilities remain that can also be tried—field delimiters, free-text, and so on.

The important principle to understand here is that these various tactics can be used singly or in combination to increase or shrink output sets. We are not so helpless as it may sometimes feel when search formulations produce results that are unacceptable for some reason. With practice, and being sure to check results by examining example records online, one can use these tactics in powerful ways to produce effective searches.

SUMMARY

Information search tactics, or moves made to further a search, are helpful in all kinds of searching. In this article their use in online searching has been described. Tactics to use early in the search and late in the search have been described and discussed. In addition, clusters of tactics to use have been recommended for cases where too many postings, too few, or no postings are produced from a search formulation.

REFERENCES

1. Marcia J. Bates. "Information Search Tactics." *Journal of the American Society for Information Science* 30:4 (July 1979) 205-214.
2. Marcia J. Bates. "Idea Tactics." *Journal of the American Society for Information Science* 30:5 (September 1979) 280-289.
3. Karen Markey and Pauline A. Cochrane. *ONTAP Online Training and Practice Manual for ERIC Data Base Searchers.* 2nd ed. Syracuse University, ERIC Clearinghouse on Information Resources, October 1981.
4. G. Carlson. *Search Strategy by Reference Librarians.* Part 3 of the Final Report on the Organization of Large Files. NSF Contract C-280. Sherman Oaks, CA: Advance Information Systems Division, Hughes Dynamics, 1964.
5. Marcia J. Bates. "The Fallacy of the Perfect Thirty-Item Online Search." *RQ* 24:1 (Fall 1984) 43-50.

THE AUTHOR

MARCIA J. BATES is Associate Professor in the Graduate School of Library and Information Science at the University of California at Los Angeles. Dr. Bates has written widely in the area of search strategy and teaches courses in online searching, science literature, information seeking behavior, and subject cataloging. Her article "Information Search Tactics" received the "Best Journal of ASIS Paper of the Year Award" for 1979. Recent publications include "Subject Access in Online Catalogs: A Design Model" (*Journal of ASIS*, November 1986), and "What Is a Reference Book?—A Theoretical and Empirical Analysis" (*RQ*, Fall 1986).

Communications to the author should be addressed to Marcia J. Bates, Associate Professor, GSLIS, University of California, Los Angeles, CA 90024; 213/825-1379.

FALSE DROPS

How They Arise. . . How To Avoid Them

by **Elaine Wagner**
Reference Librarian
Price Gilbert Memorial Library
Georgia Institute of Technology

In a discussion of online retrieval errors, Charles T. Meadow states that: "It is obviously possible for records to be relevant to the query, as stated, but not to the true requirements of the requestor."[1] Online information retrieval is accomplished by term matching, that is, searching a file for occurrences of the words requested in the search statement. "The computer searches for words not concepts."[2] Term matching of text words will always result in citations containing the requested terms, but not necessarily in the desired context (e.g., using computers to design also retrieves the design of computers).

False drop is defined as "an unwanted reference which does not pertain to the subject."[3] The term originated in the early history of document storage and retrieval systems when coded documents were represented by keypunched cards. Card sorters separated cards by dropping those representing relevant documents into specified pockets. A card representing an irrelevant document was considered a false drop. The reason for the appearance of an inappropriate reference is not always immediately apparent, but a careful examination of the complete citation will answer the user's question: "How did we get this?" Potential problem areas in text word searches are *reverse concepts, homographs, truncation, acronyms,* and *negatives.*

As end-users become end-user searchers, they will need instruction in the use of online information retrieval systems, including awareness of possible causes of retrieval errors. Many scientists and engineers, who recognize that online searching is a valuable research tool, have had experience in using the print equivalent of such databases as **COMPENDEX**, **INSPEC**, **MATHFILE**, and **NTIS**; therefore, they know something of the subject headings and classification schemes used. In pre-search interviews with intermediary searchers, faculty members and graduate students have learned Boolean logic and fundamental search strategies. Training course for end-users are usually brief and concentrate on basic online bibliographic searching. However, a few minutes spent discussing common causes of false drops, will help the trainee to anticipate potential problems.

ADVANTAGES AND PITFALLS OF TEXT WORD SEARCHES. EXAMPLES OF FALSE DROPS.

One of the advantages of an online search is the capability of retrieving information on subjects which are not index terms; it is difficult and time consuming to search a printed index under broad subject headings for a term which is too new or too specific to be included in the subject indexes. An online text word search offers the additional advantages of the syntax of natural language and greater flexibility; however, there is also an increased possibility of the inclusion of irrelevant citations. "Relevance assessments are value judgments placed on documents by individuals with information needs,"[4] and most searches, with the possible exception of those that are extremely narrow

►

.8.34

and combine several terms, do retrieve citations of only peripheral interest in addition to some irrelevant material or false drops.

Reverse concept retrieval errors are rarely anticipated, as the user is thinking only in terms of his subject interests. The words "school" and "library," truncated and limited to the same field, will retrieve both school libraries and library schools. Searches for the effects of A on B will also pick up citations on the effects of B on A. Dog bites man or man bites dog; the same words are there, but the meaning is quite different. The use of adjacency operators (**ADJ** in BRS or (**w**) in DIALOG) will retrieve the desired phrase, but will probably fail to retrieve some relevant citations. **Man ADJ bites ADJ dog**, would not pick up man bites large dog; however, the use of (**1W**) in DIALOG does allow for intervening words.

Before entering a free text term in a search statement, it is important to determine what fields are included in the basic index and possibly qualify the terms to be searched. A word from the corporate source or author field combined with a text word may yield unexpected results. Among citations on English castles, one might find an article written by authors named English or Castle. The requested terms are there but not in the desired context. Using **WITH** instead of **AND** or qualifying by field will refine the search statement.

Homographs, words with the same spelling but different meanings, can result in false drops. The computer can select citations containing the word POLISH but cannot differentiate between Polish as the language of Poland and Polish as the act or process of polishing. LEAD, the metallic element, entered in a search statement can also retrieve LEAD, to conduct or direct.

An attempt to find articles on cross burning using the search strategy: **cross and burn**; in the Wilsonline **Readers' Guide** file, retrieved the following title: "A Fire Burns Fiercely Within Him" (Olympic Cross Country Racer B. Koch) (Figure 1). The requested terms were present, but the resulting record was a false drop.

FIGURE 1

Citation retrieved from the WILSONLINE RDG file.

```
1    (READERS GUIDE)
A fire burns fiercely within him (Olympic cross country racer B. Koch)
Moore, Kenny:1943-
il pors
Sports Illustrated 60:38-40+  F 6 '84
Biography
Language: eng
Subject heading: Ski racing
Koch, Bill
Special refs: Koch, Bill
Special refs: Cross country ski racing
BRDG84006074
840211
Article
```

A search of DIALOG's file 47, **Magazine Index**, for articles on auto racing in France, using the search statement: **ss (auto or automobile) and (race or racing) and (France or French)**, retrieved 27 citations; 2 of those were about Bill France, an official of NASCAR, the U.S. National Association for Stock Car Auto Racing.

Injudicious truncation, especially truncation of a short word stem, will frequently give surprise results. **Car$1** will retrieve not only car and cars but also card, care, Carl, carp, cart, plus fifteen or more acronyms.

Acronyms can also retrieve unwanted references even when matched with terms expected to limit the search to relevant citations. A search for programming language Ada, although not an acronym, retrieved ADA-American Dental Association, etc. although Ada was combined with computer-programming or programming languages. SDI can mean Strategic Defense Initiative or Selective Dissemination of Information. Before searching for an acronym, it is wise to check an acronyms dictionary, and consider the possibility of false drops.

EXCLUSION OF RELEVANT DOCUMENTS USING THE 'NOT' OPERATOR. INCLUSION OF IRRELEVANT DOCUMENTS BECAUSE OF NEGATIVE WORDS IN THE ABSTRACT.

Online information retrieval requires a representation of the user's question as a search formulation, which combines and manipulates terms in order to retrieve relevant information in the form of bibliographic citations, while excluding material not relevant to the user's interest. One pitfall which is stressed in training online searchers is the exclusion of relevant documents by the use of the **NOT** operator. The Boolean expression, **X NOT Y**, in addition to excluding the documents about Y, results in the exclusion of those documents which discuss both X and Y. A comparison of X and Y could be valuable, but will be lost.

Although there is not always a clear division between relevant and irrelevant documents (there are usually some of marginal interest), a citation retrieved as a result of term matching with a concept which the author of the abstract states that he is *not* discussing, is an example of a record which is relevant to the search strategy but usually not to the user's interest. Some authors mention in their abstracts that certain aspects of the subject have been excluded from discussion. In reading the entire abstract, the author's intent is quite clear. However, the concept excluded by the author may be the one desired by the user. The presence of a term in an abstract is assumed to indicate what the article is about, but it may also say what the article is not about. Robert Collison, author of *Abstracts and Abstracting Services*, warns that it is "important to avoid using the negative. This, in computer searching, could result in false drops."[5] A search in BRS COMPENDEX for protection of bridges from abrasion, using: `bridge$ AND (protect$3 or prevent$3) AND abrasion`, as text words, retrieved three citations; in one entitled "Concrete Sealers for the Protection of Bridge Structures," the abstract includes the statement, "The investigation was aimed at all bridge surfaces *except* the top surface of the bridge deck which is subjected to tire abrasion." (Figure 2).

FIGURE 2

Citation retrieved from the BRS COMP file.

```
      1
AN  EI 8207-057515.
AU  Pfeifer-D-W. Scali-M-J.
IN  Wiss, Janney, Elstner & Assoc Inc, Northbrook, Ill, USA.
TI  CONCRETE SEALERS FOR PROTECTION OF BRIDGE
    STRUCTURES.
SO  Natl Coop Highw Res Program Rep 244 Dec 1981 138 p.
MJ  CONCRETE-REINFORCEMENTS.
MN  Corrosion.
XR  CONCRETE-CONSTRUCTION:    Reinforced-Concrete.
    BRIDGES-CONCRETE: Corrosion. CORROSION-PROTECTION.
IS  0077-5614.
CC  A401.   A405.   A412.   A539.
CD  NCHRB8.
AB  The objective of the investigation described was to study the
    effectiveness of different chemical surface sealers when ap-
    plied on concrete and cracked reinforced concrete which are
    subjected to different environmental conditions. Their effec-
    tiveness was established by determining if these chemical
    materials could minimize or prevent the intrusion of saltwater
    into concrete during four different laboratory test phases, in-
    cluding 24 weeks of accelerated northern and southern climate
    weathering tests. The investigation was aimed at all bridge sur-
    faces except the top surface of the bridge deck which is sub-
    jected to tire abrasion. 86 refs.
LG  EN.
```

The search statement: `aquifer$1 WITH Long ADJ Island`, was entered in the BRS NTIS database, resulting in six hits; the title of one article is "Survey of Selected Organic Compounds in Aquifers of New York State, *excluding* Long Island," (Figure 3) and a search of BRS **INSP** for: `logical ADJ inference`, produced seven citations; one abstract contains the statement: "certain operations on relations (other than logical inference) are discussed." (Figure 4). In other examples, the desired term "central control," was preceded by *no*; most of these were discussions of Ethernet, which has no central control. Although in some cases the article was of peripheral interest, the use of the negative did give a misleading idea of the content of the document. This problem can ususally be avoided by using descriptors or a combination of natural language and controlled vocabulary.

FIGURE 3

Citation retrieved from the BRS NTIS file.

```
      1
AN  PB82-139957.    8200.
AU  SCHROEDER-ROY-A.    SNAVELY-DEBORAH-S.
IN  GEOLOGICAL SURVEY, ALBANY, NY. WATER RESOURCES DIV.
    012627001.
TI  SURVEY OF SELECTED ORGANIC COMPOUNDS IN AQUIFERS
    OF NEW YORK STATE, EXCLUDING LONG ISLAND.
NT  WATER-RESOURCES INVESTIGATIONS (FINAL) 1978-80, 67
    PAGES PREPARED IN COOPERATION WITH NEW YORK STATE
    DEPT. OF HEALTH, ALBANY.
YR  1981.
JN  U8207.
RN  USGS-WRD-81-098.   USGS-WRI-81-47.
PR  NTIS PRICES: PC A04/MF A01.
CC  13B.   8H.   7C.   68D.   48G.   99A.
MJ  GROUND-WATER. ORGANIC-WASTES. WATER-POLLUTION.
MN  WATER-WELLS. SAMPLING. AQUIFERS. CHEMICAL-ANAL-
    YSIS. PHENOLS. HALOHYDROCARBONS. BENZENE. CHLO-
    RINE-ORGANIC-COMPOUNDS. PHTHALATES. GAS-CHROMA-
    TOGRAPHY. TABLES-DATA. SURVEYS. NEW-YORK.
ID  NTISDIGSWR.
AB  SAMPLES FROM 56 WELLS AT 49 SITES IN NEW YORK STATE,
    EXCLUDING LONG ISLAND, WERE ANALYZED BY GAS CHRO-
    MATOGRAPHY/MASS SPECTROMETRY FOR THE PRESENCE
    OF ORGANIC COMPOUNDS DESIGNATED 'PRIORITY POLLU-
    TANTS' BY THE U.S. ENVIRONMENTAL PROTECTION AGENCY.
    MOST SAMPLES WERE TAKEN FROM PUBLIC-SUPPLY WELLS
    TAPPING SHALLOW, PERMEABLE AQUIFERS, THE MOST SUS-
    CEPTIBLE TO CONTAMINATION. ANALYTICAL SENSITIVITY RE-
    PORTED BY THE LABORATORY FOR MOST COMPOUNDS WAS
    LESS THAN 1 MICROGRAM PER LITER, BUT CONTAMINATION
    DURING COLLECTION, SHIPPING, AND LABORATORY PRO-
    CESSING REQUIRED THAT CONCENTRATIONS BE ABOUT 10
    MICROGRAMS PER LITER BEFORE THE PRESENCE OF A COM-
    POUND COULD BE CONFIRMED. ONLY A SMALL PERCENTAGE
    OF WELLS SAMPLED WAS FOUND TO BE CONTAMINATED.
```

FIGURE 4

Citation retrieved from the BRS INSP file.

```
      1
AN  C83021727.
AU  CODD, E. F.
IN  IBM RES. LAB. SAN JOSE, CA, USA.
TI  A RELATIONAL MODEL OF DATA FOR LARGE SHARED DATA
    BANKS.
SO  COMMUN. ACM (USA). VOL.26, NO.1. 64-9. 10.
CD  CACMA2.
LG  EN. .
YR  83.
PT  J.
TC  P.
DE  DATABASE-MANAGEMENT-SYSTEMS, DATA-STRUCTURES.
ID  DBMS, SECURITY, DATA INTEGRITY, PREDICATE CALCULUS,
    RELATIONAL MODEL, LARGE SHARED DATA BANKS, PROMPT-
    ING SERVICE, APPLICATION PROGRAMS, QUERY, UPDATE,
    REPORT TRAFFIC, STORED INFORMATION, FORMATTED DATA
    SYSTEMS, TREE-STRUCTURED FILES, NETWORK MODELS,
    DATA BASE RELATIONS, UNIVERSAL DATA SUBLANGUAGE.
AB  FUTURE USERS OF LARGE DATA BANKS MUST BE PROTECTED
    FROM HAVING TO KNOW HOW THE DATA IS ORGANIZED IN
    THE MACHINE (THE INTERNAL REPRESENTATION). A
    PROMPTING SERVICE WHICH SUPPLIES SUCH INFORMATION
    IS NOT A SATISFACTORY SOLUTION. ACTIVITIES OF USERS
    OF TERMINALS AND MOST APPLICATION PROGRAMS
    SHOULD REMAIN UNAFFECTED WHEN THE INTERNAL REP-
```

►

RESENTATION OF DATA IS CHANGED AND EVEN WHEN SOME ASPECTS OF THE EXTERNAL REPRESENTATION ARE CHANGED. CHANGES IN DATA REPRESENTATION WILL OFTEN BE NEEDED AS A RESULT OF CHANGES IN QUERY, UPDATE, AND REPORT TRAFFIC AND NATURAL GROWTH IN THE TYPES OF STORED INFORMATION. EXISTING NONINFERENTIAL, FORMATTED DATA SYSTEMS PROVIDE USERS WITH TREE-STRUCTURED FILES OR SLIGHTLY MORE GENERAL NETWORK MODELS OF THE DATA. IN SECTION 1, INADEQUACIES OF THESE MODELS ARE DISCUSSED. A MODEL BASED ON N-ARY RELATIONS, A NORMAL FORM FOR DATA BASE RELATIONS, AND THE CONCEPT OF A UNIVERSAL DATA SUBLANGUAGE ARE INTRODUCED. IN SECTION 2, CERTAIN OPERATIONS ON RELATIONS (OTHER THAN LOGICAL INFERENCE) ARE DISCUSSED AND APPLIED TO THE PROBLEMS OF REDUNDANCY AND CONSISTENCY IN THE USER'S MODEL.

CC C6160, C6120.

CONCLUSION

Using only controlled vocabulary eliminates many causes of false drops; however, information on new or very narrow topics can sometimes be retrieved only through a text word search. False drops usually constitute only a small percentage of the retrieved citations, and some retrieval errors are unavoidable, especially in a search which is broad enough to achieve high recall. It is important to be aware of the reasons for the retrieval of irrelevant citations, to avoid them when possible, and to be able to explain the presence of the unwanted record to the user. "Computers are tools, not creative thinkers, and the results of an online search reflect what is literally requested, not perhaps, what is wanted."[6]

REFERENCES

1. Charles T. Meadow, *The Analysis of Information Systems*, 2d ed. (Los Angeles: Melville Publishing Co., 1973), p. 170.
2. Arthur L. Adams, "Planning Search Strategies for Maximum Retrieval from Bibliographic Databases," *Online Review* 3:373-379 (Dec. 1979).
3. Charles J. Sippl and Roger J. Sippl, *Computer Dictionary and Handbook*, 3d ed. (Indianapolis: Howard W. Sams & Co., Inc., 1980), p.197.
4. *Encyclopedia of Library and Information Sciences*, s.v. "Evaluation of Information Retrieval Systems.
5. Robert Collison, *Abstracts and Abstracting Services*, (Santa Barbara, California; ABC Clio, Inc., 1971), p. 16
6. Jane I. Thesing, "Online Searching in Perspective: Advantages and Limitations," in Maloney, ed., *Online Searching Technique and Management*, p. 19.

Communications to the author should be addressed to Elaine Wagner, Reference Department, Price Gilbert Memorial Library, Georgia Institute of Technology, Atlanta, GA 30332.

GOOD IN-HOUSE DATABASE SOFTWARE IS KNOWN BY THE COMPANY IT KEEPS

Allen, Matkins, Leck, Gamble & Mallory • Aluminum Company of America • American Film Institute, National Center for Film and Video Preservation • Bank of Boston • Baseline Inc. • The Bureau of National Affairs, Inc. • C.G. Jung Institute, Archive for Research in Archetypal Symbolism • The Catholic University of America, National Rehabilitation Information Center • Critical Inventory Program for Art on Film, a joint venture between The Metropolitan Museum of Art and The J. Paul Getty Trust • Database Services, Inc. • Deere & Company • E.P. Goldschmidt & Company, Ltd. • Engineering Information Inc. • The Getty Center for the History of Art and the Humanities Photo Archive • The Getty Art History Information Program Provenance Index • Hoffmann-LaRoche Inc. • InterAmerica Research Associates, Inc. • League of Minnesota Cities • Library of Congress, Federal Research Division • Loma Linda University Libraries • Los Angeles County Public Library, C.A.L.L. (Community Access Library Line) • M. Lee Smith Publishers and Printers • Merchant Marine Academy, The National Maritime Research Center • Michigan State University Libraries • Micromedia Limited • National Agricultural Library • National Bureau of Standards, Center for Fire Research • The National Research Council, Transportation Research Information Service • New York State Education Department • The Ohio State University • Oxford University Press • Parker Chapin Flattau & Klimpl • PennWell Publishing Company • sociological abstracts, inc. • Stepan Company • Tennessee Valley Authority, Technical Library • U.S. Borax and Chemical Corporation • U.S. Department of Agriculture, Plant Variety Protection Office • Union Carbide Corporation • University of California, Hastings College of Law Library • University of Maryland, Department of Botany . . . and so on

If you need a multi-user online system for in-house database management and information retrieval, there's an easy way to decide what's worth looking at

FOLLOW THE LEADERS... You'll be in very good company with STAR

CUADRA ASSOCIATES, INC.
2001 Wilshire Blvd., Suite 305, Santa Monica, CA 90403
Telex: 755814 CUADRA SNM Telephone: (213) 829-9972

With sales representatives and demonstration facilities in Washington, D.C., New York, Montreal, Toronto, London, and Munich

ZERO RESULT SEARCHES. . . HOW TO MINIMIZE THEM

by Steve Atkinson
State University of New York at Albany

BACKGROUND

The examination of online search strategies used in bibliographic searches has emerged as a growing field of study in the past ten years. Considerable work has been undertaken on the research of users of online catalogs and bibliographic databases on information systems such as BRS, DIALOG, and OCLC. As early as 1978 Markey, Atherton, and Newton studied the ways in which ERIC searchers used free text and controlled vocabulary statements in their searches. (1) Most recently Tolle and Hah analyzed online search patterns on the National Library of Medicine CATLINE database. (2) The method of studying search strategy has ranged from interviews with searchers to machine or computer analysis of searching sessions. The importance of these studies to the field of information science in the area of database searching lies in their application to the development of end-user, transparent, expert, front end, natural language, or artificial intelligence systems.

PURPOSE

While the two previously cited studies analyze search patterns in an online catalog and an online database, this study from the State University of New York at Albany focuses on search strategies labelled zero results by search intermediaries. The purpose of this article is to present findings based on manual observation of 187 completed zero result searches and to make recommendations to searchers who find themselves in the middle of such zero result searches.

For purposes of this article, a zero result search is defined as a yield of no relevant citations from one or more databases after a search strategy has been performed by an intermediary with the client present at the session (See statistical note*). Zero result searches represent 7% of the 2,573 searches performed at SUNYA between March 1984 and March 1985. A monthly breakdown of the statistics follows in Figure 1. Based on the 187 zero result searches, this article addresses three types of problems that can produce zero result searches: problems of multiple database searching, problems of topic negotiation, and problems of database availability. The article concludes with a summary of conceptual and practical suggestions for searchers dealing with these problems.

NOTE

The University Libraries at the State University of New York at Albany provide access to databases from BRS, BRS/AFTER DARK, DIALOG, WESTLAW, and CAS Online through the Computer Search Service unit. At the searching site in the Main Library completed search strategies are filed with their respective search request forms. Monthly statistics are compiled by a member of the support staff by inputing data from the completed request forms into a Statistical Package in the Social Sciences (SPSS) Program on the campus mainframe computer. Searches are bundled together at the end of the month and kept at least one year for the purpose of future reference in case the patron has further questions about the search or the topic is subsequently searched in additional databases. The records of strategies and searches are readily available if a searcher needs to examine the way in which the request was handled previously.

►

FIGURE 1

SUMMARY OF STATISTICS

Month Year	NUMBER OF SEARCHES	NUMBER OF ZERO RESULTS	PERCENTAGE OF TOTAL
March 1984	294	20	7%
April 1984	179	4	2%
May 1984	101	10	10%
June 1984	108	3	3%
July 1984	219	24	11%
August 1984	70	3	4%
September 1984	153	9	6%
October 1984	297	22	7%
November 1984	251	23	9%
December 1984	136	10	7%
January 1985	108	9	8%
February 1985	323	25	8%
March 1985	334	25	7%
TOTAL	2573	187	

*Statistical Note – A search is defined by our unit for statistical purposes as any request run against a single database. A search request run on ERIC, ECER, and SSCI, for example, is counted as three unique searches. A single topic run in the four files of MEDLARS on BRS (MESH, MS78, MS74, and MS70) is likewise counted as four separate searches.

MULTIPLE DATABASE SEARCHING AND FREE TEXT SEARCH STRATEGIES

The majority of the 187 zero result searches were produced from searches of multiple databases for the same topic. A user request for citations on "natural childbirth," for instance, was searched on four BRS databases, **PsycINFO**, **MEDLARS**, **Social SciSEARCH**, and **Sociological Abstracts**. After inspection by the searcher and client, citations were printed from each file except Sociological Abstracts, which yielded zero results. Another multiple database search concerned "market share, sales, promotion, or advertising for racquetball rackets." The Predicasts **F&S Index** (PTSI) database on BRS provided relevant citations, while **Management Contents** (MGMT) and Predicast **Promt** (PTSP) provided zero results. These cases support the similar findings of Evans, who observes that the needs of database users can be satisfied by a relatively small number of databases. (3) In an attempt to find as many citations as possible on a particular topic, searchers sometimes consult too many databases, some of which, not surprisingly, produce zero results.

In some cases, the zero result search was caused by searching the backfiles of a database such as Social SciSEARCH (backfile SSCB on BRS) or MEDLARS (backfiles MS78, MS74, and MS72). A request for "facial plastic surgery for people with Downs Syndrome" on MEDLARS yielded relevant results on the current file (MESH 1979+) but zero results on the first backfile (ME78, 1975-1978). In databases in which the file is divided by years, new concepts or topics appear in the current file but in many cases do not appear in the backfiles.

In some instances the searcher might have benefited from a free text search approach, increasing recall only when controlled vocabulary terms were utilized. The **TERM** database, created by Sara Knapp on BRS, allows the searcher to tap free text terms to assist in search strategy development. (4) This technique is especially useful when a control term applies only to a portion of the online database. A search on "artificial intelligence" on the **ERIC** database, for example, would require a free text approach for citations previous to 1971. The following search example demonstrates how to broaden a search on the controlled term, "artificial intelligence," from ERIC by employing TERM.

Search Example

```
TERM APR 1985
BRS SEARCH MODE – ENTER QUERY
     1 _ :     ARTIFICAL ADJ INTELLIGENCE
          RESULT        1 DOCUMENT
     2 _ :     ..P 1 TI/1
          1
TI   ARTIFICIAL INTELLIGENCE
END OF DOCUMENT

END OF DOCUMENTS IN LIST
BRS SEARCH MODE – ENTER QUERY
     2 _ :     ..P 1 ALL/1
TI   ARTIFICIAL INTELLIGENCE
ER   ARTIFICIAL-INTELLIGENCE(1971+)
```

```
ME   CONSIDER:
       ANDING terms from each column
     COMPUTERS.       INTELLIGENCE.
PS   ARTIFICIAL-INTELLIGENCE(1982+)
FT   ARTIFICIAL INTELLIGENCE. AI. FORMALIZATION OF KNOWLEDGE. COMPUTER
     SIMULATION OF INTELLIGENCE. THINKING COMPUTER. COMPUTATIONAL
     MODELS OF COGNITION. SIMULATED INTELLIGENCE. SYNTHESIZING COGNI-
     TION. INTELLIGENT SYSTEMS. DECISION SIMULATION. KNOWLEDGE ENGI-
     NEERING. KNOWLEDGE SYSTEMS. KNOWLEDGE REPRESENTATION. KNOWL-
     EDGE SYSTEM. KNOWLEDGE BASE. HEURISTICS. INFERENCE TECHNIQUE.
     DECISION MAKING SYSTEMS. EXPERT SYSTEMS. INFERENTIAL ENGINE.
     EMULATE REASONING.
```

TERM is also useful for searchers who have a free text or natural language term and want to discover either other free text terms or controlled vocabulary.

```
TERM APR 1985
BRS SEARCH MODE – ENTER QUERY
     1 _ :     MICROCOMPUTERS$1
        RESULT        3 DOCUMENTS
     2 _ :      ..P 1 TI/1-3

       1
TI   COMPUTER LITERACY.

       2
TI   MICROCOMPUTERS.

       3
TI   COMPUTERS.

END OF DOCUMENTS IN LIST
BRS SEARCH MODE – ENTER QUERY
     2 _ :      ..P 1 ALL/2

       2
TI   MICROCOMPUTERS.
ER   MICROCOMPUTER(1980+).
ME   MICROCOMPUTERS*.
FT   MICROCOMPUTER. MICROCOMPUTERS. MICROPROCESSORS. PERSONAL
     COMPUTING. PERSONAL COMPUTERS. MICROS. WORD PROCESSORS. CON-
     SIDER NAMES OF SPECIFIC MICROCOMPUTERS. HOME COMPUTERS. MICRO
     COMPUTERS. PORTABLE COMPUTERS. DESKTOP COMPUTERS. HANDHELD
     COMPUTERS. APPLE I. APPLE II. APPLE III. CP/M. ATARI. TRS-80. VIC 20.
     IBM PERSONAL. MICROPROCESSOR CHIPS. DRAGON. CBM. DBASE II. PEACH-
     TREE. MICROPRO. WORDSTAR. VISICORP.
```

In addition to the online database to assist in term selection, there are databank cross files, **CROS**, **DIALINDEX**, and **DATA BASE INDEX** which assist in database selection before the term selection process. These helpful searchaids should be utilized previous to multiple database searches to ensure that the databases selected for searching have postings on the key concepts. CROS on BRS, DIALINDEX on DIALOG, and Data Base Index on SDC offer search assistance before the searcher approaches the terminal with a multifile search. A quick search of the most important key terms from the client's request will save both time spent in search strategy preparation and money spent in multiple files. A list of articles discussing multifile searching and presearch helpers follows.

FIGURE 2

Multiple Database Searching and Presearch Helpers

General Articles:

Hawkins, Donald T. "Multiple Database Searching: Techniques and Pitfalls." *ONLINE* 2 (April 1978):9-15.

Wanger, Judith. "Mutliple Database Use." *ONLINE* 1 (October 1977):35-41.

Databank Specific:

Antony, Arthur. "The Database Index." *DATABASE* 2 (December 1979):28-33.

Dolan, Donna R. "The BRS CROSS Database." *DATABASE* 3 (December 1980): 50-55.

Databank Comparisons:

Blair, John. "CROSS-Database Searching A Chemical Compound: Comparing Lockheed, SDC, BRS, and NLM." *ONLINE* 5 (April 1981):46-61.

In summary, searchers should begin a multifile search by entering an abbreviated search (key concepts) on the databank cross searching files (CROS, DIALINDEX and DATA BASE INDEX) and if necessary utilize TERM on BRS or the database thesaurus to add controlled vocabulary or free text terms to build the search strategy. Although the success of the search depends in large part on the interview, maximizing the use of these presearch techniques minimizes the probability of zero results.

TOPIC NEGOTIATION: VAGUE VS. SPECIFIC REQUESTS

Searchers can readily predict in some search interviews that the user request will produce low postings or zero results. At one extreme, when the client has read or knows little or nothing about the topic, a vague request can pose considerable difficulties for the searcher, who must try to construct a search strategy from his or her own knowledge and the insufficient criteria provided by the client. In these cases, the client commonly changes his or her own thinking about the topic in response to the citations discovered. Such vagueness on the part of the client may lead to zero results because it precludes the formulation of a successful search strategy.

At the other extreme, overly specific user requests can likewise contribute to difficulties in the development of a search strategy. A faculty member's request for citations from the MEDLARS database on "early damage to the hypocampus and its effect on behavior at maturity in the rat" exemplifies this problem. The client added further restrictions to the search strategy during the interview, such as damage to newborn or infant rats, and he also wanted to look at studies that analyzed follow-up behaviors of adult populations. In requests of this type, it is the responsibility of the searcher to build a strategy that includes the most relevant concepts from the interview. Entering all of these criteria would produce low postings and ►

probably zero results. A diagram of the search restrictions or concepts in this search request follows in Figure 3.

FIGURE 3

SEARCH DIAGRAM

Population	Population modification	Technique	Target	Outcome/ measurement
rat rats	infant newborn	damage by radiation or lesions	hypocampus or fornix or septal nuclei or septum pellicudum	behavior or feeding or mazes

The selection of the most relevant terms with input from the client and use of the thesaurus resulted in the following strategy run on MEDLARS and its backfiles. This is an example of how to work with the user who demands that all of his or her criteria be included in the search.

Search Example

```
MESH 1979 – JUN 1985
BRS SEARCH MODE – ENTER QUERY
    1 _ :        HIPPOCAMPUS.SJ,SN.
      RESULT        4370 DOCUMENTS
    2 _ :        SEPTAL-NUCLEI
      RESULT         488 DOCUMENTS
    3 _ :        SEPTUM-POELLUCIDUM
      RESULT         618 DOCUMENTS
    4 _ :        FORNIX.TI,AB.
      RESULT         342 DOCUMENTS
    5 _ :        1  2  3  4
      RESULT        5225 DOCUMENTS
    6 _ :        DAMAGE$1 WITH (LESION$1 RADIATION).TI,AB.
      RESULT        1661 DOCUMENTS
    7 _ :        5 AND 6
      RESULT          35 DOCUMENTS
    8 _ :        DAMAGE$1.TI.
      RESULT        4620 DOCUMENTS
    9 _ :        5 AND 8
      RESULT          61 DOCUMENTS
    10 _ :        7 OR 9
      RESULT          93 DOCUMENTS
    11 _ :         ..P 10 TI,DE/1-4

       1
TI   Amnesia following basal forebrain lesions.
DE   AMNESIA: et. BASAL-GANGLIA. LIMBIC-SYSTEM. ADULT. AMENSIA: pa, px,
     ra. BASAL-GANGLIA: pa, ra. CASE-REPORT. FEMALE. HIPPOCAMPUS: pa, ra.
     HUMAN. LIMBIC-SYSTEM: pa, ra, MALE. MIDDLE-AGE. PERSON-
     ALITY-DISORDERS: et, pa, px. PSYCHOLOGICAL-TESTS. SUPPORT-U-S-GOVT-
     P-H-S. TOMOGRAPHY-X-RAY-COMPUTED.

       2
TI   Multimodal amnesic syndrome following bilateral temporal and basal forebrain
     damage.
DE   AMNESIA: et. BASAL-GANGLIA. LIMBIC-SYSTEM. TEMPORAL-LOBE. AMNESIA:
     pa,px,ra. BASAL-GANGLIA: pa. CASE-REPORT. DIENCEPHALON: pa. HIP-
     POCAMPUS: pa. HUMAN. LIMBIC-SYSTEM: pa. MALE. MIDDLE-AGE. PERSON-
     ALITY-DISORDERS: et, pa, px. PSYCHOLOGICAL-TESTS. SUPPORT-U-S-GOVT-
     P-H-S. SYNDROME. TEMPORAL-LOBE: pa.

       3
TI.  Usefulness of parenteral kainic acid as a model of temporal lobe epilepsy.
DE   DISEASE-MODELS-ANIMAL. EPILEPSY-TEMPORAL-LOBE. KAINIC-ACID. PYR-
     ROLIDINES. AMYGDALOID-BODY: me. ANIMAL. BRAIN-DAMAGE-CHRONIC:
     cl. EPILEPSY-TEMPORAL-LOBE: cl,me. GLUCOSE: me. HIPPOCAMPUS: me.
     RATS. SYNDROME.

       4
TI   Maturation of kainic acid seizure-brain damage syndrome in the rat. III. Post-
     natal development of kainic acid binding sites in the limbic system.
DE   BRAIN-DAMAGE-CHRONIC: cl. EPILEPSY: cl. KAINIC-ACID: me. LIMBIC-SYS-
     TEM: me. PYRROLIDINES: me. AGE-FACTORS. ANIMAL. AUTORADIOGRAPHY.
     BINDING-SITES. BRAIN-DAMAGE-CHRONIC: me. EPILEPSY: me. HIP-
     POCAMPUS: me. RATS. SUPPORT-NON-U-S-GOVT.
```

Search Results

Of the four databases searched on MEDLARS for this topic, the following results were found by recycling the most relevant subject headings and subheadings discovered from the 93 documents found in the initial search.

MESH (1979-present)
93 documents browsed
15 documents chosen as relevant

MS74 (1971-1974)
29 documents browsed
9 documents chosen as relevant

MS78 (1975-1978)
35 documents browsed
13 documents chosen as relevant

MS70 (1966-1970)
11 documents browsed
0 documents chosen as relevant

1200/2400 Baud Searching or Downloading

Printing titles and subject headings from all of the documents retrieved from the search in the previous example at 1200 baud allows for the reformulation of the strategy. It also permits screening of all documents to assist in selecting the most relevant. The rapid change in communication speeds of modems and terminals to online services has changed the way in which topics can now be searched. The expansion from 30 characters per second a few years ago to 1200/2400 baud and the capability to download on microcomputers allows for a few options not previously feasible either technically or financially.

BOOLEAN OPERATOR "AND"

When two or three concept columns are ANDed (Boolean Operator) in a search strategy and low postings or zero results occur, the search can be broken apart. Printing titles and subject headings of 100 records or more online or downloaded to a microcomputer may prove to be cost effective on databases with low to medium connect hour rates ($25-$50 per hour or $.42/$.83 per minute), and no charges for browsing titles and subject headings. This technique enables the searcher to identify the relevant citations from the search and its reformulation and print (and pay) only for the citations that he or she feels are important and useful.

This searching technique is especially useful when a search strategy ends in low postings or zero results from the use of the "AND" operator.

As an example, a search for "teaching methods for language acquisition in autistic children" provided only a few citations on the ERIC database. Combining only two of these concepts (broadening the request), "teaching methods for language acquisition" provides relevant citations after the searcher and the client browse and select documents regardless of the population.

The same strategy can be used for "language acquisition with autistic children." The searcher can sample at 1200/2400 baud or by downloading from the ERIC database and select the most relevant citations.

Venn diagrams and a sample search showing this technique searching at 1200 baud is provided.

language acquisition

teaching techniques

autistic children

language acquisition

autistic children

teaching methods

autistic children

Search Example

```
ERIC 1966—APR 1985 (BOTH)
BRS SEARCH MODE—ENTER QUERY
     1 _ :      AUTISM.MJ,MN.
       RESULT        774 DOCUMENTS
     2 _ :      TEACHING-METHODS
       RESULT      44516 DOCUMENTS
     3 _ :      CLASSROOM-TECHNIQUES
       RESULT       4356 DOCUMENTS
     4 _ :      EDUCATIONAL-STRATEGIES
       RESULT       2856 DOCUMENTS
     5 _ :      TEACHING-MODELS
       RESULT       1065 DOCUMENTS
     6 _ :      2  3  4  5
       RESULT      50069 DOCUMENTS
     7 _ :      LANGUAGE-ACQUISITION
       RESULT       6504 DOCUMENTS
     8 _ :      LANGUAGE-SKILLS
       RESULT       4685 DOCUMENTS
     9 _ :      ORAL-LANGUAGE
       RESULT        597 DOCUMENTS
    10 _ :        7  8  9
       RESULT      10993 DOCUMENTS
    11 _ :        1 AND 6 AND 10
       RESULT         12 DOCUMENTS
    12 _ :        1 AND 6
       RESULT         65 DOCUMENTS
    13 _ :        1 AND 10
       RESULT         91 DOCUMENTS
```

Future studies into searching behavior should focus on the effect of the rate of transmission of data or downloading to microcomputers on the searching performance and behavior of both end-users of online systems and intermediaries.

DATABASE AVAILABILITY

A third type of problem arises when searches are requested on topics in fields for which the major print indexes, abstracts, full-text journals, collections, or library catalogs in a discipline are not available online as databases. A case in point is the field of anthropology. Brody and Lambert point out that none of the printed indexes in the area of anthropology are available online, so a creative approach must be taken to choose the most appropriate file(s). (5) Multiple databases such as Social SciSEARCH, Sociological Abstracts, or PsycINFO may cover this area peripherally. If the request is tied to some other discipline such as medicine or language, it can be searched in files such as Excerpta Medica, MEDLARS, or Language and Language Behavior Abstracts. In these cases, the lack of online resources in the field necessitates use of database indexes such as CROS, DIALINDEX, and Data Base Index. A search for "machine translation of Mayan hieroglyphs," for example, retrieved relevant citations from the INSPEC database after many of the social science files provided zero results. The searcher discovered postings that led to relevant citations only by checking CROS on BRS.

There are no online equivalents to the major print indexes, abstracts, and collections in the area of anthropology. The only database listed in reference database directories is the French database, Francis: Ethnologie, included in Cuadra's *Directory of Online Databases.* This file, however, is not available on any of the major American systems at this time.

The following list of Anthropological resources illustrates the problem; although there are numerous print sources, the searcher must employ alternative databases to search an anthropological topic online.

Indexes and Abstracts:

Anthropological Literature: *An Index to Periodical Articles and Essays.* South Salem, N.Y., Redgrave Publishing Company., 1979-. Quarterly.

Museum of Mankind. Library. *Anthropological Index to Current Periodicals in the Museum of Mankind Library* (incorporating the former Royal Anthropological Institute Library). 1977-. Quarterly.

Royal Anthropological Institute of Great Britain and Ireland Library. *Index to Current Periodicals Received in the Royal Anthropological Institute.* London, 1962-. Quarterly.

International Bibliography of Social and Cultural Anthropology. 1955-. Prepared by the International Committee for Social Sciences Documentation in Cooperation with the International Congress of Anthropological and Ethnological Sciences. N.Y.: Tavistock Publications, 1958-. Annual.

►

Levine, Morton H. *A Topical Guide to Volumes 1-21 (New Series) of The American Journal of Physical Anthropology*. Philadelphia, Wistar Institute Press, 1971.

Abstracts in Anthropology. Farmingdale, N.Y.: Baywood Publishing Co., 1970-. Quarterly.

Annual Review of Anthropology. Palo Alto, CA, Annual Reviews Inc., 1972-. Annual.

Library Catalogs:

Harvard University. Library. Geography and Anthropology: Classification Schedules; Classified Listing by Call Number; Chronological Listing; Author and Title Listing. Cambridge, Published by Harvard University Library; distributed by Harvard University Press, 1979.

Harvard University. Peabody Museum of Archaeology and Ethnology. Library. Catalogue: Authors. Supplement 3. Boston, G.K. Hall, 1975.

Harvard University. Peabody Museum of Archaeology and Ethnology Library. Catalogue: Authors, Boston, G.K. Hall, 1963-. Catalogue: Subjects. Boston, G.K. Hall, 1963-.

Smithsonian Institution. Department of Anthropology. National Museum of Natural History. Catalog to Manuscripts at the National Anthropological Archives. Boston: G.K. Hall, 1975.

Collections:

HRAF Files. New Haven, CT: Human Relations Area Files Inc., 1958-.

Journals:

American Journal of Physical Anthropology. N.Y.: Alan R. Liss, Inc., 1918-. Monthly.

Database:

FRANCIS: ETHNOLOGIE
Bibliographic
Producer: Centre National de la Recherche Scientifique, Centre de Documentation Sciences Humaines (CNRS/CDSH) France.
1972+
Online Service: Centre National de La Recherche Scientifique, Centre de Documentation Sciences Humaines, and Télésystèmes-Questel
Content: Worldwide literature on ethnology and social anthropology
Language: French with keywords in English and French

From: Cuadra Associates
Directory of Online Databases
V. 5 # 3 Spring 1984
Santa Monica, CA

SUGGESTIONS FOR SEARCHERS

As for suggestions for searchers who find themselves in a search strategy which produces zero results, I recommend some of the tactics Marcia Bates defines in her article, "Idea Tactics." (6) The summary table of her article including the tactics and their definitions follows. I have added sample search behaviors from my day to day searching.

Summary of Idea Tactics and Definitions

TACTIC	DEFINITION
Think	To stop and think about the search and try to come up with new ideas for searching difficulties
Brainstorm	To generate new ideas and to suspend critical reactions until the ideas are well-formed and can be fully evaluated
Meditate	To develop and utilize the skill of quickly achieving a mind state in which rational and intuitive capacities work together in solving search problems
Consult	To ask a colleague for suggestions or information in dealing with a search
Rescue	To check for possibly productive paths still untried, in an otherwise unproductive approach
Wander	To move among one's resources, being receptive to alternative sources and new search ideas
Catch	To catch oneself in an unproductive search, to notice that a change in approach might be more productive
Break	To break a habitual search pattern, that is, put it aside temporarily, in order to take a search tact more suited to the particular problem in question

Breach	To breach the boundaries of one's concept of the limits of the intellectual or physical territory in which one searches to respond to a query
Reframe	To examine implicit frame-of-reference information coming along with the user's query, to deal with it explicitly with the user, and to negotiate changes in the query to eliminate distortions
Notice	To watch for clues that revise one's notions of the nature of the question or of the answering information
Jolt	To think "laterally," to snap out of conventional thought patterns and redefine a problem in a dramatically different way
Change	To change something, anything, in one's search behavior—to try a different source, a different term, a different subject field, etc.
Focus	To look at the query more narrowly, in one or both senses: (1) to move from the whole query to a part of it or (2) to move from a broader to a narrower conceptualization of the query
Dilate	To look at the query more broadly, in one or both of the senses: (1) to move from a part of the query to the whole query or (2) to move from a narrower to a broader conceptualization of the query
Skip	To shift laterally in one's view of the query, in one or both of two senses: (1) to move from searching one part of a complex, multipart query to another one or (2) to view the query from, and search on, another angle, that angle being neither broader or narrower, but simply different
Stop	To cease searching temporarily on a problem and do something else

RECOMMENDATIONS

Think	Use good problem solving skills, define problem or question, ask questions, list the most important concepts, ask yourself what you already know about this topic, use CROS, DIALINDEX, or Database Index to select possible databases, use foresight to predict probable search results
Brainstorm	Look for new ideas, check thesauri, make associations, write down all possible subject headings and free text terms associated with each concept, use the TERM database, use rotated displays, use free text terms with controlled vocabulary when designing the search strategy
Meditate	Relax, contemplate, use intuition, follow instinct
Consult	Go for assistance, (two heads are better than one), find someone who knows more about the database, call the database producer's 800 toll free number, consult someone who knows more about the subject area
Rescue	Try an alternative approach, look for an ideal title, look for ideal concepts to appear together in the same sentence of the title or abstract paragraphs, analyze subject headings from citations browsed for reformulation
Wander	Analyze search results for possible ideas including discovered terms, codes, descriptors, identifiers, and reformulate, use your own knowledge and experience with a topic
Catch	Be aware of possible changes that need to be made, realize when search results are not satisfactory

▶

Break	Don't use the same search pattern for every search, i.e., controlled vocabulary only, try cited references, free text, or citation pearling, print citations from search results that yield positive results for reformulation
Breach	Ask open ended questions
Reframe	Place the search question in a new context either during the interview or after the search has been entered and observed on the database, adjust the search strategy and search results to meet the needs of the client
Notice	Watch for free text terms, controlled vocabulary, identifiers or codes to reformulate for an alternative strategy from the search results, elicit responses to the search results from the client, observe his or her responses to the search results
Jolt	Change search terms, approach subject from a different perspective
Change	Try something different from new terms to new databases
Focus	Break the request down to some of its parts, think about the request in a narrower sense, think about the request by analyzing its parts rather than the whole
Dialate	Think about the request more broadly, look at a search as a whole rather than its parts
Skip	Search part of a problem then search other parts of the problem
Stop	Stop searching, take a break, off continue, go for lunch, go for a swim, go for a walk, sleep on it, try again tomorrow, use hindsight

CONCLUSION

Known item searches for author, title, author/title, or call number on bibliographic databases present few problems to end-users compared to subject, identifier, code, or free text (title, title/abstract, title, abstract, full text) searches which usually require the services of an intermediary. In order to dispense with the intermediary – who performs functions like interviewing, vendor selection, database selection, search strategy development, selection of vocabulary, and performing, analyzing, and reformulating searches – end-user systems will have to incorporate many of the tactics defined by Bates in their design. Some of the most noteworthy of these tactics, are "dilating," "refining," "skipping," "changing," and "wandering," all of which are particularly important to the success of searches which might otherwise yield "zero results."

REFERENCES

1. Markey, Karen; Atherton, Pauline; and Newton, Claudia. "An Analysis of Controlled Vocabulary and Free Text Search Statements in Online Searches." *Online Review* 4 (September 1980): 225-236.
2. Tolle, John E. and Hah, Sechang. "Online Search Patterns: NLM Catline Database." *Journal of the American Society for Information Science* 36 (March 1985): 82-93.
3. Evans, John E. "Database Selection in an Academic Library: Are All Those Big Multi-File Searches Really Necessary?" *ONLINE* 4 (April 1980): 35-43.
4. Knapp, Sara D. "Creating BRS/TERM, A Vocabulary Database for Searchers." *DATABASE* 7 (December 1984): 70-75.
5. Brody, Fern and Lambert, Maureen. "Alternative Databases for Anthropology Searching." *DATABASE* 7 (February 1984): 28-33.
6. Bates, Marcia J. "Idea Tactics." *Journal of the American Society for Information Science* 30 (September 1979): 280-289.

Communications to the author should be addressed to Steve Atkinson, State University of New York at Albany, Reference Services, 1400 Washington Ave., Albany, NY 12222, telephone (518) 457-3347.

GRAPHING ONLINE SEARCHES WITH LOTUS 1-2-3

by Olle Persson
INFORSK
Sociologiska institutionen, Umea universitet
Sweden

In a recent article in *ONLINE* (March '85), Alberta Bailey described how Lotus 1-2-3 can be used to create a card file database on a microcomputer (1). Here, I will show how Lotus 1-2-3 can be used to create graphs using downloaded online searches as raw material. I will present information on most commands applied and the readers should be able to make replications of their own.

One of the great advantages of the Lotus software is its capability to import text files directly into the worksheet. When this is done, one can easily design graphs. Three steps are required.

1. DOWNLOADING

Most communication software allows downloading of online searches. Commands and references can be saved on a separate file. For this demonstration, I have downloaded a table from the **PTS/PROMT** database showing the robot production by country (Table 1). The receiving file must have the extension ".prn" in order to be imported into the Lotus worksheet (for example "table.prn").

2. IMPORTING THE DOWNLOADING FILE INTO THE WORKSHEET

The file called "table.prn" can now be imported to the 1-2-3 worksheet.

a. Ideally, before importing the "table.prn" file, it should be edited. Lotus makes a distinction between labels and numbers. Therefore, each label must be enclosed in quotes, e.g. "W Germany." This can be done with a common package such as WordStar. When all labels are marked in this manner, the file can be imported to the worksheet with the "file import number" command.

b. There is no absolute need to edit the ".prn" file before importing it to Lotus. The editing could be made using the worksheet commands instead. Place the cursor at row 1 and column A – the home position. When issuing the "file-import-text" command, each row of the file will be inserted as labels beginning at the top of the worksheet. Note that the labels now contain both text and numbers. Therefore, the next step will be to edit each label and delete the numbers.

When this is done we need to fill in the numbers. Move to the top of column H and issue the "file-import-number" command. This puts all numbers of the "table.prn" file to the right of the label column. Now the worksheet is prepared for graphing. Don't worry if there are some superficial numbers left at the top or bottom of the worksheet.

3. MAKING GRAPHS

The graph menu offers almost an indefinite number of ways to design a graph. Go to the graph menu (/Graph). Enter the column "W Germany. . .Sweden" as the X-range, 3.5. . .1.3 as the A-range and 4.8. . .1.9 as the B-range. Name

▶

the axes and titles of the graph with the Options-Titles command and prepare labels for the A- and B-range (Options-Labels command). Issue the Type-Bar command and then the View command—et voila! (Figure 1).

This little exercise demonstrates the possibility of using conventional software for microcomputers to refine online searches. A large number of numerical databases are now available, and the use of such databases could be enhanced by micros and graphics software such as Lotus 1-2-3.

4. BIBLIOMETRICS

A common question among librarians is "which are the most important journals within field x?" One way of answering such a question, is to study the distribution of articles by journals. In the database service of ESA/IRS, there is a special command that automatically produces rank frequency lists of words or codes for most fields of the references—the zoom command. For a given set of items, it is possible to receive a rank list of the most productive authors, organizations, or journals. To illustrate how this command works, I logged onto the **INSPEC** database of ESA/IRS and created a set of all articles from 1984 containing the words "information retrieval" and "online." The zoom command was applied to the journal name field of these references to produce the rank frequency list shown in Table 2.

Congratulations to the editors of the Online, Inc. journals.

Now, this list of most productive journals can be directly imported to the Lotus worksheet as described in section 2 b. The graph in Figure 2 shows a rather typical Bradford distribution with the most productive journals on the left hand side. Note that the y-axis is a cumulation of the rank frequency list (articles) and that the X-range (the number of journals) has been created by the data-fill command of the worksheet. Also note that the journal names have been inserted directly into the graph by using the journal labels as they were imported into the worksheet (use the Options-Data-label command).

To sum up, Lotus 1-2-3 is a useful software for graphing data retrieved by online searches. In my opinion the procedures described in this article are very easy to apply. I'm also convinced that the end-users of online services will appreciate the result of such an exercise.

REFERENCES

1. Bailey, Alberta S., "Creating and Maintaining a Database/Databank Comparison System with Lotus 1-2-3 On a (Most-of-the-time) IBM Compatible Micro," *ONLINE* 9 (March '85)2, 86-92.

TABLE 1

Downloaded search of a table from PTS/PROMT

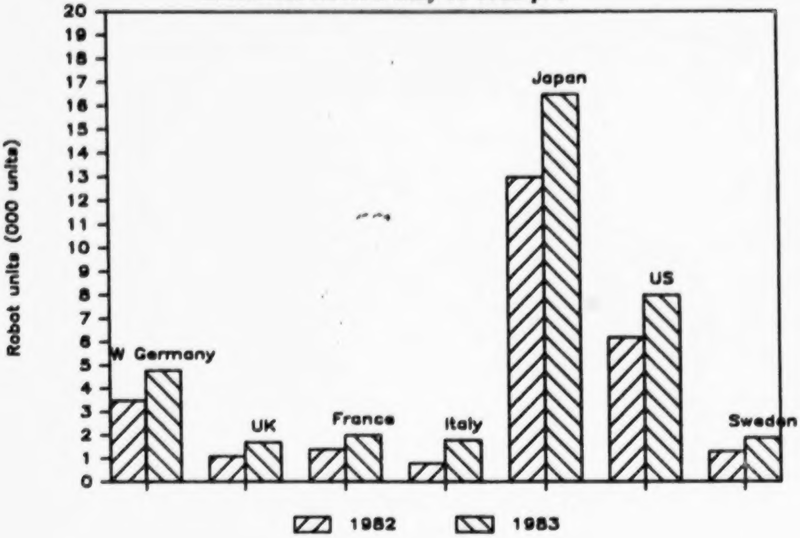
```
File16:PROMT – 72-85/Aug (UD = 8508W1)
(Copr. PREDICASTS Inc. 1985)

         Set      Items        Description

? srobot?;ssf = table;c1*2
          1        3259        ROBOT?
          2       75524        SF = TABLE
          3         102        1*2
? t3/5/1
3/5/1
1214309
  Estimated robot production/. World: Robot output (000 units).
Electrical Review   May 10, 1985   p. 6

                       1982        1983
W Germany               3.5         4.8
UK                      1.1         1.7
France                  1.4         2.0
Italy                   0.8         1.8
Japan                  13.0        16.5
US                      6.2         8.0
Sweden                  1.3         1.9

*0123 *Western World *3569400 *Industrial Robots *62 *production data
4WGE West Germany 3569400 Industrial Robots 62 production data
4UK United Kingdom 3569400 Industrial Robots 62 production data
4FRA France 3569400 Industrial Robots 62 production data
4ITA Italy 3569400 Industrial Robots 62 production data
9JPN Japan 3569400 Industrial Robots 62 production data
1USA United States 3569400 Industrial Robots 62 production data
5SWE Sweden 3569400 Industrial Robots 62 production data
```

TABLE 2

A zoom of journal names. Articles on information retrieval online, accession year 1984(INSPEC via ESA/IRS)

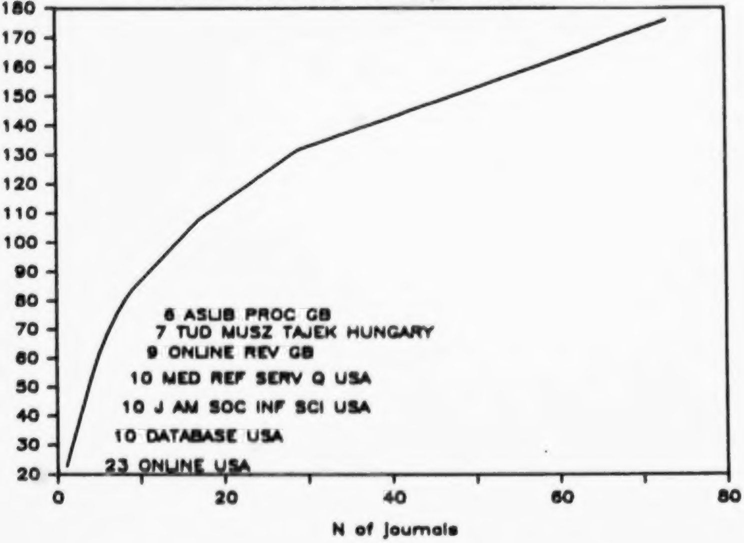
```
?zoom5(176)jo     (zoom on set 5:"online information retrieval")

                          Text Analysis Results
      Frq         Words/Phrases

       23         ONLINE USA
       10         DATABASE USA
       10         J AM SOC INF SC I USA
       10         MED REF SERV Q USA
        9         ONLINE REV GB
        7         TUD MUSZ TAJEK HUNGARY
        6         ASLIB PROC GB
        5         INF SERV USE NETHERLANDS
        4         LASIE AUSTRALIA
        3         CESK INF TEOR PRAXE CZECHOSLOVAKIA
        3         INF TECHNOL LIBR USA
        3         INF TECHNOL RES DEV GB
        3         J CHEM INF COMPUT SC I USA
        3         J INF IMAGE MANAGE USA
        3         J INF SC I PRINC PRACT NETHERLANDS
        3         TIDSKR DOK SWEDEN
        3         VINE GB
        •
        •           (Additional rows with frequency 2 or 1)
        •
```

FIGURE 1
Lotus 1-2-3 graph of Table 1



ESTIMATED ROBOT PRODUCTION
Electrical Review May 10 1985 p. 6

FIGURE 2
Lotus 1-2-3 graph of Table 2



SCATTER OF ONLINE IR—ARTICLES IN 1984
Online search of INSPEC (zoom command)

Communications to the author should be addressed to Dr. Olle Persson, INFORSK, Sociologiska institutionen, Umea universitet, 901 87 Umea, Sweden.

A PRIMER ON CROSSFILE SEARCHING IN THE INFORMATION ACCESS COMPANY DATABASES

by Judith Sovner-Ribbler
President
Searchline Associates, Inc.
Brookline, Massachusetts

Each of the databases produced by Information Access Company (IAC) is unique in its scope and coverage. As a collection of databases, they form a complementary library of information which supports research in the areas of business and industry, current affairs, consumer interests, and the law.

BUSINESS, MARKETING, PRODUCT, AND INDUSTRY INFORMATION

NEWSEARCH, **TRADE AND INDUSTRY INDEX** (which includes the Area Business Databank subfile and PR Newswire), **TRADE AND INDUSTRY ASAP**, **MANAGEMENT CONTENTS**, and **NATIONAL NEWSPAPER INDEX** allow the user to:

- obtain company profiles including financial, management, marketing, and advertising data
- research the activities of competitive companies including Fortune 500 companies and emerging companies
- track new product developments and evaluations
- research the market share of a product or the market size of an industry
- keep up-to-date in management theory, organizational behavior, and administration
- search for private companies and regional business developments

THE BUSINESS AND THEORY OF TECHNOLOGY

Research the business of technology, searching NEWSEARCH, **COMPUTER DATABASE**, TRADE AND INDUSTRY INDEX, and TRADE AND INDUSTRY ASAP to:

- monitor industry news for high technology systems, electronics, engineering, communications, computer hardware and software
- track technology companies, products, industries and people
- find evaluations of hardware and software products for professional and personal computer users
- research the use of computers and automation in any subject area
- locate information on a new concept, buzz word, jargon, fact or statistic

CURRENT AFFAIRS AND PEOPLE IN THE PRESS

Use NEWSEARCH, **MAGAZINE INDEX**, **MAGAZINE ASAP**, NATIONAL NEWSPAPER INDEX, and TRADE AND INDUSTRY INDEX to:

- locate interviews, biographies, and obituaries of celebrities, politicians, authors, writers, business executives, world leaders, and other people in the news
- summarize and analyze current issues and events from 1959 to the present
- track political campaign developments, medical discoveries, world events and foreign affairs

THE LAW

NEWSEARCH and **LEGAL RESOURCE INDEX** look at the legal perspective of a topic, and complement sources that provide the complete text of a case or a statute. These IAC databases allow the user to:

- retrieve a state-of-the-art review or analysis of a particular case or a statute
- track news on legislation, regulations, trials, malpractice, product recalls, and liability issues
- research people cited in the legal literature
- find articles and evaluations on law office automation

CONSUMER AFFAIRS AND PRODUCT EVALUATIONS

Use NEWSEARCH, MAGAZINE INDEX, MAGAZINE ASAP, and NATIONAL NEWSPAPER INDEX to:

- locate reviews of restaurants, movies, records, theatre, books, compact disks, and other entertainment
- provide descriptions, evaluations, and comparisons of consumer products
- identify opinions on social and political issues

RELATIONSHIP BETWEEN IAC DATABASES

NEWSEARCH serves as a repository for the most recent indexing for all of the more than 1900 journals received at IAC. Each journal article is indexed to a specific database, and loaded into the NEWSEARCH file. For example, all articles selected from *Business Week* are indexed to MAGAZINE INDEX, but loaded initially into NEWSEARCH. These and other MAGAZINE INDEX records are coded in the banner line of the record as **DATABASE: MI File 47**, and may be retrieved from NEWSEARCH using the controlled prefix field **SF=MI?** or **SF=MI File 47**.

The seven subfiles that compose NEWSEARCH may be searched individually or in combination to retrieve the latest additions to IAC databases.

The seven subfiles that compose NEWSEARCH may be searched individually or in combination to retrieve the latest additions to IAC databases. For example **SS SF=LRI?** limits the search in NEWSEARCH to LEGAL RESOURCE INDEX records. **SS SF=(TI? or CDB)** will limit a search to COMPUTER DATABASE and TRADE AND INDUSTRY INDEX.

Subfiles may also be searched in LEGAL RESOURCE INDEX, and TRADE AND INDUSTRY INDEX/ASAP. **SS SF=ABD** limits a search to Area Business Databank records within TRADE AND INDUSTRY INDEX. **SS APARTHEID NOT SF=NNI?** will eliminate duplicate records from NATIONAL NEWSPAPER INDEX when crossfile searching in LEGAL RESOURCE INDEX.

See Table 1 for the complete list of subfiles within IAC databases.

UPDATING IAC DATABASES

Records added to NEWSEARCH during the previous calendar month are removed on the weekend following the thirteenth of each month. The removed records then form the monthly updates for NATIONAL NEWSPAPER INDEX, MAGAZINE INDEX, TRADE AND INDUSTRY INDEX, LEGAL RESOURCE

▶

INDEX and MANAGEMENT CONTENTS, and a semimonthly update for COMPUTER DATABASE. The full-text portion of the records for MAGAZINE ASAP and TRADE AND INDUSTRY ASAP are attached to the index record and loaded into the full-text files at this time.

TABLE 1

RELATIONSHIP BETWEEN IAC DATABASES

DATABASE	SUBFILE (SF=)
NEWSEARCH	MI?,NNI?,LRI?,TI?,ABD, CDB
NATIONAL NEWSPAPER INDEX	NNI?
MAGAZINE INDEX	MI?
MAGAZINE ASAP	MI? (ASAP only)
TRADE AND INDUSTRY INDEX	TI?, ABD, MI?, NNI?, LRI?
TRADE AND INDUSTRY ASAP	TI?, MI? (ASAP only)
LEGAL RESOURCE INDEX	LRI?, MI?, NNI?, TI?
MANAGEMENT CONTENTS	MC?
COMPUTER DATABASE	CD?

Note: The PR Newswire is considered a journal title (JN=PR NEWSWIRE).

After the NEWSEARCH purge, a record will be placed in more than one IAC database, if the indexers judge that its subject is relevant to the contents of each file. For example, selected newspaper articles about industry or legal issues on a topic may be added to both TRADE AND INDUSTRY INDEX and LEGAL RESOURCE INDEX.

CROSSFILE SEARCHING ON DIALOG

IAC databases provide articles on a topic from the perspective of newspapers, magazines, business reports, legal analysis, popular opinion, and interviews. Together these different views may produce an in-depth profile of a single subject.

To perform a crossfile search, develop the core of the search strategy in NEWSEARCH to find the most current information. Then execute the strategy across other IAC files. Narrow topics usually lend themselves to crossfile searching with a minimum of revisions in strategy, while broad or complex topics require that the strategy be modified to focus the search to the subject contents of the file. These modifications are facilitated in the IAC databases since the same subject terms and access fields are used across files, and duplicate records can be eliminated through subfiles. See Table 2 for a summary of DIALOG Access Fields for IAC databases.

LOTUS DEVELOPMENT CORPORATION: A CROSSFILE STUDY

Assembling a full profile of Lotus Development Corporation from information found in the IAC files will aptly illustrate the power of crossfile searching, utilizing subject terms, common access points, and subfiles.

Begin the search in NEWSEARCH to track the current activities at Lotus. For example, an investment firm searched the daily press for reports about Lotus when they were "going public," and monitored the public's reaction to Mitch Kapor's departure from operations.

```
File 211:NEWSEARCH WK=05
(Copr.1987 IAC)

Set      Items   Description

?ss lotus()development/co;end/savetemp

S1         64    LOTUS/CO
S2        380    DEVELOPMENT/CO
S3         61    LOTUS()DEVELOPMENT/CO
Temp Search-save "TB051" stored

 3/5/1
591322   DATABASE: MI File 47
Home software kaboings again. (personal computer market rises
from the dead)
Fortune v115 p10(1) Feb 16, 1987
CODEN: FORTA
Illustration; photograph
COMPANY NAME(S): Lotus Development Corp.–cases; Broderbund
Software Inc.–finance; Electronic Arts Inc.–finance; Mosaic Software
Inc.–cases; Paperback Software International–cases
DESCRIPTORS: Computer software industry–finance
```

Search Tips:

1. To search for articles that discuss a company, use the company name field /CO or CO=. Truncate or expand the prefix field CO= for best results.

```
E CO=LOTUS
SS CO=LOTUS DEVELOPMENT CORP?
```

2. Use the temporary search command for transfer to another file.

```
SS CO=LOTUS DEVELOPMENT CORP?
SS MITCH(S)KAPOR/NA
END/SAVETEMP
```

Next search NATIONAL NEWSPAPER INDEX or MAGAZINE INDEX. Because these databases do not utilize subfiles, searching these databases first will simplify the elimination of duplicate records later in the search.

TABLE 2

SUMMARY OF DIALOG ACCESS FIELDS

BASIC INDEX

FIELD	SUFFIX	INDEXING	FILES	NOTES
Abstract	/AB	Word	CD,MC,N/S	
Caption	/CP	Word	All	
Company	/CO	Word	All	
Descriptor	/DE	Word/Phrase	All	
Named people	/NA	Word	All	
Product name	/PN	Word	All	
Text	/TX	Word	ASAP	Displays in MI,TI
Title	/TI	Word	All	

ADDITIONAL INDEXES

FIELD	PREFIX	INDEXING	FILES	NOTES
Article type	AT=	Word/Phrase	All	
Author	AU=	Phrase	All	
Availability	AV=	Phrase	All	
Case citation	CC=	Word/Phrase	LRI,N/S	
Case name	CN=	Word/Phrase	LRI,N/S	
Coden	CO=	Phrase	All	
Column inches	CI=	Phrase/Range	All	Newspapers only
Company	CO=	Phrase	All	
Edition	ED=	Phrase	All	Newspapers only
Geographic code	GC=	Phrase	All	
Geographic location	GN=	Word/Phrase	All	
Grade	GR=	Phrase	All	Reviews only
Journal name	JN=	Phrase	All	
Jurisdiction	JU=	Word/Phrase	LRI,N/S	
Length indicator	LI=	Phrase/Range	ASAP	
Named person	NA=	Phrase	All	
Operating system	OS=	Phrase	CD,N/S	
Product name	PN=	Phrase	All	
Program language	PL=	Phrase	CD,N/S	
Publication date	PD=	Phrase/Range	All	
Publication year	PY=	Phrase/Range	All	
Publisher	PU=	Word	LRI,N/S	Legal book reviews only
SIC code	SC=	Phrase	All	
Special feature	SF=	Phrase	All	
Statute cite	SO=	Phrase	LRI,N/S	
Statute name	ST=	Word/Phrase	All	
Subfile	SF=	Phrase	All	
Ticker symbol	TS=	Phrase	All	
Update	UD=	Phrase/Range	All	
Week	WK=	Phrase/Range	N/S	

Summary of DIALOG Access Fields is reprinted with the permission of Information Access Company from *ACCESS TO ACCESS: AN ONLINE USER'S GUIDE TO IAC DATABASES*. 3rd edition. Belmont, CA. Information Access Company. 1986. page 49.

► 47

Execute the saved search from NEWSEARCH in NATIONAL NEWSPAPER INDEX to locate newspaper articles from five major newspapers: the *New York Times, Wall Street Journal, Christian Science Monitor, Los Angeles Times*, and *Washington Post*. All stories of national, international, and business interest are indexed.

```
File 111:NATIONAL NEWSPAPER INDEX – 79-87/JAN
(COPR. 1987 IAC)
Set     Items   Description
?exs tb051
S1        117   LOTUS/CO
S2       1047   DEVELOPMENT/CO
S3        102   LOTUS()DEVELOPMENT/CO
?ss s3 and mergers/de
          102   S3
S4      41288   MERGERS/DE,CO
S5         15   S3 AND MERGERS/DE,CO

 5/5/1
1231441   DATABASE: NNI File 111
Lotus completes acquisition. (Graphic Communications merger)
Wall Street Journal p6(E) June 26 1986
CODEN: WSJOAF
col 6 001 col in.
EDITION: Thu
COMPANY NAME(S): Lotus Development Corp. – acquisitions and
mergers; Graphic Communications Inc. – acquisitions and mergers
```

> **A subdivision used in conjunction with the company or named person field qualifies the context in which the company or business executive is being discussed.**

Search Tip:

1. A subdivision used in conjunction with the company or named person field qualifies the context in which the company or business executive is being discussed. The use of subdivisions such as management, finance, officials and mergers provides the means to refine a search.

```
SS LOTUS()DEVELOPMENT/CO AND MERGERS/CO,DE
SS LOTUS()DEVELOPMENT/CO AND OFFICIALS/CO
```

Next use MAGAZINE INDEX to locate an interview or biography of Mitch Kapor.

```
File 47:MAGAZINE INDEX -1959-MARCH 1970, 1973-86/SEP
(COPR. IAC) FMT 9 = $3.50
Set      Items   Description
?ss mitch?(s)kapor/na and at=biography and av=full text
S1        1774   MITCH?/NA
S2          50   KAPOR/NA
S3          49   MITCH?/NA(S)KAPOR/NA
S4       90357   AV=FULL TEXT
S5          15   MITCH?(S)KAPOR/NA AND AV=FULL TEXT

 5/5/2
2085854   DATABASE: MI File 47 *Use Format 9 for FULL TEXT*
Mitch Kapor steps down from position as chairman of Lotus.
(Lotus Development Corp.)
Duffy, Richard
PC Week v3 p1(2) July 15 1986
AVAILABILITY: FULL TEXT Online
LINE COUNT: 00057
SIC CODE: 7372
NAMED PEOPLE: Kapor, Mitchell – resignation, retirement, etc.;
Manzi, Jim P. – selection and appointment
COMPANY NAME(S): Lotus Development Corp. – officials and
employees
DESCRIPTORS: Computer program industry – officials and employees
```

Search Tips:

1. Several IAC fields make use of a controlled vocabulary. For example, there are eleven different article types and each one is a controlled vocabulary term (i.e. biography, editorial, evaluation, etc.). An article type combined with a topic will retrieve evaluations (of products), or biographies (of individuals).

```
SS MITCH(S)KAPOR AND AT=BIOGRAPHY
SS PN=LOTUS AND AT=EVALUATION
```

2. Print the full text of an article in MAGAZINE INDEX only if the banner line reads **Use Format 9 for Full Text**. To identify only full-text articles for online delivery in File 47, combine the search results with **AV=FULL TEXT**.

Search TRADE AND INDUSTRY INDEX to retrieve articles relating to corporate strategy, industry analysis, joint ventures, contracts, earnings, marketing and advertising. PR Newswire and Area Business Databank will be searched automatically as part of TRADE AND INDUSTRY INDEX. Selected business-related articles from other IAC files are added to TRADE AND INDUSTRY INDEX.

```
File 148:TRADE AND INDUSTRY INDEX 81-87/JAN
(COPR. 1987 IAC) * FMT 9 = $3.50
Set      Items  Description
?exs tb051
S1          349 LOTUS/CO
S2         3048 DEVELOPMENT/CO
S3          313 LOTUS()DEVELOPMENT/CO
?ss (s3 and sc=7372) not sf=(mi? or nni?)
            313 S3
S4         9832 SC=7372
S5       279531 SF=MI?
S6       529567 SF=NNI?
S7           62 (S3 AND SC=7372) NOT SF=MI? OR NNI?)

 7/5/1
814635   DATABASE: TI File 148
CAE packs lure business software cos. (computer aided engineering)
Cain, Matthew
Electronic News v32 p36(1) Dec 1 1986
SIC CODE: 7372
COMPANY NAME(S): Lotus Development Corp. Engineering and Scien-
tific Products Div. – product introduction; Cullinet Software Inc. –
product introduction
PRODUCT NAME(S): Lotus Measure (computer program) – product in-
troduction; Lotus Manuscript (computer program) – product introduction
DESCRIPTORS: Computer-aided engineering – computer programs;
```

```
Microcomputers – computer programs; Computer software industry –
product introduction
```

Search Tips:

1. Add the article type, **COMPANY PROFILE**, to limit your search to stories with detailed company information.

```
SS LOTUS()DEVELOPMENT/CO AND AT=COMPANY PROFILE
```

2. To eliminate articles previously retrieved in NATIONAL NEWSPAPER INDEX and MAGAZINE INDEX, use the Boolean operator **not** with the subfiles you wish to withdraw.

```
SS S5 NOT SF=(MI? OR NNI?)
```

3. To identify a company's activity in a specific industry, combine the company name with the SIC code or the descriptor term for the industry. SIC codes are assigned to 4 digits, but may be truncated after the second digit.

```
SS LOTUS/CO AND SC=7372
SS LOTUS/CO AND COMPUTER SOFTWARE INDUSTRY
```

Search LEGAL RESOURCE INDEX to retrieve articles on legal issues related to software protection.

```
File 150:LEGAL RESOURCE INDEX – 80-86 DEC
(COPR. 1986 IAC)

Set      Items   Description

ss software and litigation/de

S1         541   SOFTWARE
S2       47526   LITIGATION/DE
S3          46   SOFTWARE AND LITIGATION/DE

 3/5/1
2092156   DATABASE: LRI File 150
Copyright protection expanded for software.
Raysman, Richard; Brown, Peter
New York Law Journal v196 p1 Sept 10 1986
col 1 050 col in.
EDITION: Wed
JURISDICTION: United States
Whelan Associates v. Jaslow Dental Laboratory Inc. 609 F.Supp. 1325
(E.D. Pa. 1985)
GEOGRAPHIC CODE: NNUS
DESCRIPTORS: Copyright – Computer programs; Computer programs –
litigation
```

Search Tips:

1. LEGAL RESOURCE INDEX uses the descriptor *LITIGATION* while other IAC databases use the more general term *CASES*.

```
SS LOTUS()DEVELOPMENT/DE AND LITIGATION /DE
```

2. If a company is a plaintiff or a defendant in a case, search the case name field **CN=** in LEGAL RESOURCE INDEX.

```
CN=LOTUS
```

3. Selected records of law-related articles are added to LEGAL RESOURCE INDEX from NATIONAL NEWSPAPER INDEX, MAGAZINE INDEX, AND TRADE AND INDUSTRY INDEX. Rather than eliminating the three IAC subfiles previously searched, combine the search strategy with **SF=LRI?**

```
SS S5 AND SF=LRI?
```

Search the full text of TRADE AND INDUSTRY ASAP to retrieve additional articles on the Lotus product, HAL.

```
File 648:Trade and Industry ASAP 83-87/Jan
(Copr. IAC 1987)

Set      Items   Description

?exs  tb051

S1        133    LOTUS/CO
S2       1011    DEVELOPMENT/CO
S3        122    LOTUS()DEVELOPMENT/CO
?ss (lotus and hal) not s3

S4       2312    LOTUS
S5       1576    HAL
          122    S3
S6         13    (LOTUS AND HAL) NOT S3

 6/5/1
1779746   DATABASE: MI File 47 *Use Format 9 for FULL TEXT*
Don't give up your ease-of-use goals for software-yet. (column)
PC Week v3 p29(1) June 10 1986
ARTICLE TYPE: column
AVAILABILITY: FULL TEXT Online
LINE COUNT: 00057
SIC CODE: 7372
DESCRIPTORS: Computer programs – purchasing; Computer program
industry – technological innovations
```

Search Tips:

1. Highly specific areas such as new or private companies, or technical jargon, require that all articles available must be retrieved. In these cases, use truncation and broad connectors such as **and** to allow for maximum retrieval. The full-text files are particularly well-suited for retrieval of any mention of a word, person, company or specific concept.

2. Search the company name free-text, and then eliminate it in the company name field to direct the search to the text of articles not indexed to Lotus Development Corp. in the corresponding IAC index files.

Search COMPUTER DATABASE for information on computer-related hardware, software, applications, company services and products. Both current and retrospective articles are indexed with abstracts for high tech industries and growing areas such as local area networks and artificial intelligence.

```
File 275:COMPUTER DATABASE 83-87/ISSUE02
(COPR. 1987 INFORMATION ACCESS COMPANY)

Set      Items   Description

?exs tb051

S1        698    LOTUS/CO
S2       1598    DEVELOPMENT/CO
S3        691    LOTUS()DEVELOPMENT/CO
?ss s3 and artificial()intelligence
          691    S3
S4       2823    ARTIFICIAL
```

►

```
S5      3321   INTELLIGENCE
S6      2715   ARTIFICIAL(W)INTELLIGENCE
S7        23   S3 AND ARTIFICIAL()INTELLIGENCE

 7/5/1
1197624
Repositioning AI. (artificial intelligence) (column)
Irsfeld, Mitch
Computer Systems News ISSUE: n292 PAGINATION: p34(2)
PUBLICATION DATE: Dec 8, 1986
ARTICLE TYPE: column
The possibility of marketing artificial intelligence (AI) technology may
mean abandoning the term AI. There is a common perception that AI,
with languages such as LISP, Prolog and SmallTalk, is only appropriate
for programming complicated and proprietary applications. Exper-
Telligence Inc. has been a leader in promoting AI for low-end or com-
mercial applications development by marketing its ExperCommon LISP
for operating on the Apple Macintosh. Lotus Development Corp.
understood the stigma associated with AI when it introduced its HAL
product, a natural-language interface for Lotus 1-2-3. Nowhere in the
marketing or promotional materials could the term AI be found, and
some analysts suggest that had Lotus called HAL an AI program, its level
of success would be much lower. One analysts suggests that incor-
porating AI in common applications will take away the mystique and
make the use of rule-based programming, voice recognition and natural
language more common. The same individual suggests that value added
resellers are the best candidates for executing this agenda.
PRODUCT NAME: Apple Macintosh (microcomputer) usage
COMPANY NAME: ExperTelligence Inc., marketing; Lotus Development
Corp., marketing TICKER SYMBOL: LOTS AAPL
SIC CODE: 7372; 3573
DESCRIPTORS: Artificial Intelligence; Market; Market Analysis; LISP;
Natural Language Interfaces; Enhancements; Outlook; Software
```

Search Tips:

1. Since May, 1986, the same access fields and article types exist in COMPUTER DATABASE as the other IAC files. Subfiles do not appear in COMPUTER DATABASE, but COMPUTER DATABASE is a subfile of NEWSEARCH.

2. The subject terms used for COMPUTER DATABASE are technical in nature, and therefore are different from those used in the other IAC files. These subject terms are available within the database as an online thesaurus.

3. COMPUTER DATABASE provides indexed records with abstracts while NATIONAL NEWSPAPER INDEX, MAGAZINE INDEX, TRADE AND INDUSTRY INDEX, and LEGAL RESOURCE INDEX are index files. TRADE AND INDUSTRY ASAP, MAGAZINE ASAP, and PR Newswire (within NEWSEARCH) are index records with full-text.

CONCLUSION

The user's manual *ACCESS TO ACCESS: An Online User's Guide to IAC Databases*, 3rd edition, published in December, 1986, brings together all the editorial policies, access points, controlled vocabularies, master lists, and thesauri used by the IAC indexers. These new aids are extremely useful for searching the nine IAC files. The following sections of *ACCESS TO ACCESS* are relevant for company and industry searches:

Part II: ACCESS POINTS
- ARTICLE TYPE for company profiles, directories
- CASE NAME for companies in litigation
- COMPANY for companies and subdivisions
- DESCRIPTOR for subject terms and subdivisions
- JOURNAL NAME for PR Newswire
- SIC CODE for Standard Industrial Classification Codes
- TICKER SYMBOL for company ticker symbols (PR Newswire only)
- TITLE for annotations and newspaper columns

Part IV: USER AIDS
- COMPANY AUTHORITY LIST for the top 500 companies
- QUICK REFERENCES for article types, database subfiles
- SIC CODES for industries

Part V: JOURNAL LISTS
- BUSINESS SUBDIVISIONS for use with companies and industries

While there are several excellent business databases on DIALOG, I have concentrated on the scope of the IAC files for company, industry, product, and personnel information. The use of subject terms and indexing for companies and industries will assist the searcher in focusing retrieval on the exact resource or specific concept needed, and the use of subfiles will eliminate the duplication of records between files during a crossfile search.

THE AUTHOR

JUDITH SOVNER-RIBBLER, President of Searchline Associates, Inc., Brookline, MA, is an information broker and online database training consultant.

Communications to the author should be addressed to Judith Sovner-Ribbler, President, Searchline Associates, Inc.; 8 Griggs Terrace; Brookline, MA 02146; 617/277-2991; DELPHI—Librarian; EasyLink—62876562; DIALMAIL—10897.

PASSWORD PROTECTION:

Some Dos And Don'ts

by M. D. Parker
Systems Consultant

INTRODUCTION

When you subscribe to a databank, generally, four pieces of information are provided:

- Telephone access number
- Network identification code for your databank
- Databank account code
- Databank password

You dial the telephone access number to connect your terminal to a network that contains several online services. You must then enter the network identification code to connect to your databank. Once connected, you must enter your account code. Your databank usually *assigns* your account code; however, it may allow you to *select your own* account code. Finally, you must enter your databank password to identify yourself as an authorized user.

Your password must remain confidential, as it is the only way your databank can identify you as an authorized user. If a thief steals your password, the system, allows the thief to use your account without your knowledge.

You must take steps to prevent theft of your databank account. Protecting your password from theft requires knowledge in the following areas:

- Password Modification
- Password Selection
- Password Protection
- Usage Verification

This article will provide information in these areas. ►

DATABANK PASSWORD MODIFICATION

The first step is to modify your password. This section will provide you with information about changing your password.

RULE 1: Change your databank password immediately after you receive it.

Changing your password prevents use of your databank account by databank employees or by a thief who may intercept this information.

Normally, your password can be changed by accessing one of the databank menu screens. Your databank documentation should include instructions to change your password. If you cannot find this information, then call the online service. The databank representative will be glad to provide you with this information.

CompuServe and The Source allow the user to change the password while online. The Source encourages that the password be changed monthly. Conversely, on all BRS services, both DIALOG services, and SDC, the password holder must write to the search service for a new password. In some cases, the account must be closed and a new account with new password opened. On BRS and SDC, the second level security password may be easily changed online as frequently as the user desires.

RULE 2: Change your databank password at least once every three months.

Frequent changes of your password prevent a thief from continuing to use your account without detection.

RULE 3: When a group of people use your databank account, change your password whenever someone leaves the group.

This protects you from an ex-group member who may continue to use your account. Your databank may allow you to save information online which may contain confidential business data. If you do not change your password, this confidential data may be compromised.

RULE 4: Change your databank password as often as you feel necessary.

This rule, though obvious, is often forgotten. For instance, you might be demonstrating the databank to a group of people and, someone from the group steals your password and uses your account. In a safe location, after such demonstrations, change your password.

PASSWORD SELECTION

Modifying your password is only one part of protecting your account. Just as important is the choice of your password. The following rules will help you to choose a good password.

RULE 5: A databank password should be at least six characters in length.

Computer security studies have found that many people choose very short databank passwords because they are easy to remember. However, such short passwords provide very little protection against theft. In fact, some passwords are so simple, that it is easy for a thief to guess. Databank passwords such as "DATA," invite theft of your account. A long password is safer from theft. However, your databank may limit the length of your password.

RULE: 6 Choose passwords that consist of two unrelated common words separated by a special character such as a comma, period, number sign, question mark, or slash.

This combination gives the safest and easiest to remember databank passwords. A password of all random characters is unsafe. This type of password must be written to remember it, thus making it easier for a thief to find. An example of an effective password is "TAPE#PASTURE." However, please do not use this example for your own password.

RULE: 7 DO NOT choose passwords that are easily associated with yourself.

Many people choose passwords that directly relate to themselves. These passwords are easily stolen. Unacceptable databank passwords include:

- names of any relatives
- addresses
- phone numbers
- birthdays
- license plate numbers
- ZIP codes
- account numbers

RULE 8: DO NOT use CONTROL characters in your databank password.

Your online service may allow the use of nonprinting CONTROL characters in your password. These characters are obtained by first pressing the CONTROL key and then pressing a letter. However, CONTROL characters perform special functions by your databank. The easiest procedure is to avoid CONTROL characters in your password. In addition, your password should not contain the ESCAPE character. Terminal manufacturers use the ESCAPE character to perform special functions. Using the ESCAPE character in your password can cause some really bizarre behavior by your terminal.

Using the ESCAPE character in your password can cause some really bizarre behavior by your terminal.

RULE 9: DO NOT choose identical databank passwords for more than one databank.

As the selection of databanks become larger, you will probably have accounts on several systems. It is tempting to use one password for all your online accounts. However, all your accounts are at risk if your single databank password is stolen.

PASSWORD PROTECTION

The protection of your databank password is just as important as the modification and selection. This section will describe several rules to prevent your password from theft.

RULE 10: Store your password in a safe place.

Your databank account is only as safe as your password. If you leave your password close to your terminal, it will be stolen and used by a thief. For instance, in the movie *War Games*, the main character was able to change his grades because he discovered the password written near a terminal. The safest procedure is to store your password away from your terminal.

RULE 11: Memorize your password.

Memorizing your password prevents it from being written, making it difficult to steal. However, be sure to select your password according to the rules described earlier.

RULE 12: DO NOT give your databank password to others.

This is an extremely important rule to remember. The host computer can separate you from a thief only by your password. If you give your password away, it is impossible for the system to block the thief. Do not provide your password over the phone, since it can be easily overheard by a thief.

Your databank may provide a utility that allows different users all over the country to talk (i.e. type) to each other. Do not provide your password while using these utilities. You should inform your online service when you see someone requesting databank passwords.

RULE 13: DO NOT store your passwords on a floppy or hard disk.

Storing your password on a disk can make it easier for you to use your databank account. However, this makes it easier for a thief to steal your password. A thief can simply copy your disk to obtain your password.

RULE 14: Enter passwords ONLY at a databank prompt.

Your online service protects your password from disclosure during the signon sequence in two different ways. One method your databank may use is a "strike over" mask consisting of random characters. Entering your password over this mask ensures that your password cannot be discovered through your terminal output. Your system may also prevent the display of your password as you enter it. This ensures that your password is not visible in your terminal output.

Your databank communicates in one of two modes: full duplex and half duplex. In the full duplex mode, input and output between you and the system are carried on two distinct paths. To see what you have sent in full duplex mode, the system must echo what you entered. Full duplex is the mode most likely in use.

The half duplex mode shares one path between you and the host computer. In other words, while receiving data output, you cannot enter data or commands to your databank. The half duplex mode is most common on TELEX systems. Whether your modem operates in full duplex or half duplex mode often determines if your password appears on your terminal as you enter it.

The following trick is helpful in protecting your password if your system operates in the half duplex mode. Before entering your password, switch your modem from half duplex mode to full duplex mode.

Your databank documentation specifies the mode you should use. The following trick is helpful in protecting your password if your system operates in the half duplex mode. Before entering your password, switch your modem from half duplex mode to full duplex mode. This prevents what you enter from being displayed on your terminal. When you enter your password it will not appear on your terminal. Switch your modem back to the half duplex mode after entering your password.

►

RULE 15: If in doubt as to the validity of a databank password prompt, hang up the phone and start over.

Your databank may allow you to use utilities provided by other users. Sometimes, these utilities may request your password. The utility will store your password for the thief who provided the utility. If you are unsure whether a password should be typed, simply hang up the phone and start over. This type of deception is called a "Trojan horse."

DATABANK ACCOUNT VERIFICATION

Although you may carefully select, change and protect your password, there is still risk of theft. The best way to deter theft of your databank account use is to compare your own databank use with the system's accounting records.

RULE 16: Record your databank use and compare it against your reported use.

Recording the following items helps in resolving unauthorized databank charges:

- date and time of login
- network and network address used
- network port number (given shortly after the initial connection to the network)
- length of your online session

Your online service may provide detailed information about how your charges were obtained. If you see unusual charges in your account, it is possible that your password has been stolen. If this occurs, immediately change your password and inform the service.

DATABANK RESPONSIBILITY

Much of the responsibility for protecting your databank account rests with you. However, the online service has responsibilities also. It should not display your password on your terminal, but should provide a dense character mask to obscure it. If your databank does not provide this service, it shows a lack of concern for your business. In this case, register a complaint.

Your online service should allow you to change your password independently of its representatives, which often results in a password different from what you requested. Unfortunately, your databank may not allow you to modify your passwords except in this manner. Fortunately, these databanks are on the decrease.

Ideally, a databank should not have direct access to your password. This can result in theft of your password despite your efforts to protect it. A databank should not modify your password over the phone. A simple phone call, by anyone, can change your password without your knowledge. Obviously, this means that you will not have access to your own account, while the thief runs up charges. Instead, the databank should mail a new password to your registered address. It is irritating to lose access to your account for a few days; however, the resulting security provided to your account is worth the wait. In addition, the online service should not accept address changes over the phone for similar reasons.

A databank should not modify your password over the phone. A simple phone call, by anyone, can change your password without your knowledge.

CONCLUSION

Your databank password is confidential information that is the key to your account. You must protect it and change it whenever you suspect it has been compromised. Although the online system provides some protection, you must take steps to protect your databank account. Your responsibility can be made easier by following appropriate techniques for password protection, selection, modification and databank account verification.

THE AUTHOR

M.D. PARKER is a systems consultant who has worked extensively in the design and development of computer software systems. He holds a Masters Degree in Computer Sciences.

Communications to the author should be addressed to Mr. M. D. Parker, MDPC, P.O. Box 261369, San Diego, CA 92126-0993, telephone (619) 695-8635.

“MEATBALL SEARCHING”

The Adversarial Approach To Online Information Retrieval

by Robert F. Jack*
NASA/University of Kentucky**
Technology Applications Program

A DAY IN THE LIFE. . .

6:00 a.m. The metallic ring of a cheap alarm clock breaks the silence at the Jack household.

6:01 a.m. Bob Jack, NASA/UK TAP's online searcher, tries to answer the clock as if it were a telephone. Annoyed at the lack of a human response, he throws the offending device under the bed and resumes his interrupted sleep.

7:10 a.m. The wailing of a baby in the next room awakens Mrs. Jack, who reminds her still-sleeping husband that he must begin the morning routine in order to get to work on time. Reluctantly, our hero arises, drinks two cups of lukewarm instant coffee, smokes three cigarettes, and showers – simultaneously. (The water from the shower extinguishes the first two cigarettes almost as soon as they are lit).

7:30 a.m. With a hardy “Heigh ho, Silver!” the online searcher takes his trusty (rusty?) auto at daredevil breakneck pace through the streets of Lexington to the main campus of the University of Kentucky, swearing and cursing at those who would impede his progress.

7:59 a.m. The online searcher has made it to the office on time, with minimal damage to the flora and fauna of the Bluegrass.

8:00 - 8:20 a.m. A quick cup of coffee with the rest of the staff, discussing the events of the day, special activities, etc.

8:20 a.m. Back to the office of the online searcher. The office suite had been a dormitory resident advisor's apartment in an earlier lifetime, and the searcher's office had been its walk-through living room. The strategic placement of bookcases now vaguely resembles a wall, much in the spirit of Edgar Allen Poe's “The Cask of Amontillado.” More coffee and more cigarettes are consumed as mail is opened, the terminal is switched on, and the excitement begins. Thus commences the typical day for a self-described “meatball searcher.” (Editor's comment: A term that derives from the “meatball” surgery as practiced on the popular TV program, M*A*S*H, e.g. a jack-of-all-trades practitioner).

STYLES OF SEARCHING

Online searching is a personal experience, reflecting a combination of skill, luck, imagination, and curiosity. There are probably as many different styles of searching as there are searchers.

There have been a number of studies on search strategy models as well as searcher technique. Oldroyd and Citroen (1977), for example, developed two general models of search strategies: Model A, in which only the most relevant concepts are selected and combined logically as an initial step; and Model B, in which very broad concepts are selected before any Boolean techniques are used [1]. Some very interesting results have been obtained from research into the role a searcher's professional experience and training play in searching behavior. Some of the work by Carol Fenichel suggests that experience and training alone do not seem to account for the differences in online behavior [2]. A rather similar study by Helen Howard noted that, in effect, there are alot of different kinds of “training;” moreover, certain “environmental” factors such as the presence of the end-user and the availability of

►

---

*This research was supported in part by the United States National Aeronautics and Space Administration and the University of Kentucky under Contract #NASW-3367. The comments and opinions expressed herein are those of the author and do not necessarily represent the views of the aforementioned agencies.

**Since writing this article, the author has changed jobs. See Communications to the author at the end of this article for current address.

database thesauri may well play some role in style [3]. Julia Rholes and Judith Droessler concluded that "searchers are characteristically more flexible and adaptive than nonsearchers, having a lesser need for things to proceed along a planned, orderly course" [4].

For the moment, we can look at most styles as informal or formal. In the informal style—by my definition, at least—the searcher simply selects the term or phrase as a free text item (perhaps performing some elementary Boolean logic with other free-text words or phrases) and displays the citations online or prints them offline. The result is usually a relatively inexpensive search with fair-to-middling recall and precision.

I think of the formal or textbook style as the time-consuming method of carefully evaluating controlled vocabulary lists, constructing the strategies offline, avoiding free-text searching, and in general, prior planning. These searches also are often rather inexpensive, although an experienced searcher would probably achieve a fairly high level of precision in such a search.

Search training by the commercial online services tends to center around the informal approach. "See, just type in the word you're looking for, and you'll get *everything* you need."

By now, the reader who has been through a few different kinds of training sessions may recognize something personal. Search training by the commercial online services tends to center around the informal approach. "See, just type in the word you're looking for, and you'll get *everything* you need." Sure, as long as the author used the buzzword in his title, or the abstractor included the term in the summary, or the buzzword is already part of the controlled vocabulary for the database. Because most online training is primarily instruction in the mechanical aspects of searching, it can be regarded, as Donna Dolan pointed out, as "left brain oriented" [5].

Formal searching brings to my mind that solid week of MEDLARS training I underwent a few years ago. For someone who doesn't have a great personal awareness of medical terminology and practice, it opens up a whole new world of technique and perception. Ever notice how MEDLARS searchers always seem to be in close proximity to a current copy of *Medical Subject Headings (MeSH)?* Have you further

If it exists, we'll find it.

Information is your business. Delivering it is ours.

No matter where you are, no matter where the document is that you need, the Information Store will get it to you. Our high-tech, worldwide network assures prompt, cost-effective delivery of reports, government documents, specifications, proceedings or anything ever published, anywhere in the world.

Don't delay in getting the information you need...415 543-4636

The Information Store
Professional Document Retrieval Services

The Information Store is a service mark of The Information Store, Inc.
Cable: Dialorder On Tyme: ORBDOC INFOSTORE Facsimile: 415/543-9147 BRS: 119; The Source: TCC 598 CompuServe: 74105.1661 MCI Mail: ID177-2317

noticed that said *MeSH*, as a result of all this use, looks like it got into a fight with the gorilla in the luggage commercial? Of course, medicine is a highly structured discipline, and its language is oriented to the very specific. On the other hand, if the indexing is so good that free text searching isn't necessary, why go through the expense of putting the database online? Why not just print cumulative indexes in hard copy or microfiche and completely forget about interactive online searching?

That, in a nutshell, is that. Thesaurus-keyword searching is for technical areas with highly developed vocabularies and/or for sophisticated searchers, while free text searching is for topical searches or those performed by lazy, sloppy searchers. Now we can split into two armed camps and call each other names.

CLIENT NEEDS

Such an outlook, unfortunately, ignores the fact that online searching is a service—a service provided to patrons, co-workers, superiors, subordinates, the general public, or whoever. Employing one technique exclusively penalizes the unusual client.

For example, consider the MEDLARS searcher in a hospital setting. The majority of clients are most likely to be physicians, many of whom were no doubt exposed to online searching in medical school. The years of education they receive has taught them the value of using a standardized vocabulary in describing medical conditions, symptoms, bacteria, all those gooey things inside the human body, and what not. The leading journals reflect this precision, and, indirectly, so do the leading databases in biomedicine. The physician is looking for a few really good citations, and is more than likely to get them if the searcher sticks with a formal approach to the online inquiry.

Another scenario: the academic library, where a university professor (in, say, a college of business administration) is researching a paper for publication. Scholarly publications place great value on refereed materials and on citing the "right" sources; so, a controlled vocabulary search in ABI/INFORM, Management Contents, or a similar database, coupled with an author/cited reference search in Social SciSearch, is likely to do just fine. Again, a fairly formal approach.

A third case: a college freshman—a political science major—wants to write a paper on the movie "The Day After" for an English composition class. He or she may be looking for popular literature that discusses how viewers responded to the television images, but cash is limited. The solution: a "down and dirty" free text search of Magazine Index, PAIS, PsycINFO, or a similar file.

As long as money is a consideration, and a client doesn't really want "everything" on any given subject, each approach can be seen to have distinct advantages, depending on the client.

SEARCHER PERSONALITY

The background and personality of the searcher is a large factor in how a search can (or should) be performed. A number of studies have tried to identify the characteristics of the online searcher; my personal favorite is Judith Wanger's "it," that characterizes some mysterious "not quite definable" quality:

> . . . the craft part of the searching process [that] represents a different quality at each step in the strategy process. . . "It" is probably the ability. . .to begin relating a search problem to the world of information. . . This "it" requires some unique translation abilities: the searcher listens to a statement and explanation of the search problem, and hears not only the key words and concepts but hears, as well, the subtle nuances of the problem that are related to certain characteristics of databases [6].

MEATBALL SEARCHING

Fans of the late television program M*A*S*H will recall that the fictionalized doctors referred to their practice as "meatball surgery." The implication was getting the job done as quickly as possible, in handling almost superhuman workloads, required a special knack. Part of this talent depends on the ability to recognize the circumstances in which expediency is more important than finesse of technique. Similarly, online searching—at least the way I do it—might be called "meatball searching."

"Meatball searching" is deliberately vague, just like the job. The effort requires generally comprehensive searches, often on unusual topics and for a wide diversified audience, and, more often than not, with tight deadlines.

"Meatball searching" is deliberately vague, just like the job. The effort requires generally comprehensive searches, often on unusual top-

►

ics and for a widely diversified audience, and, more often than not, with tight deadlines. The subject areas are frequently multidisciplinary in scope, requiring searches of several databases offering distinctive treatment of the topic at hand. In many cases our clients are not technical experts in the fields to which the search relates, and they may lack institutional access to the information they need to address specific questions.

The "meatball searcher" has to identify the databases most likely to produce "usable" results; more often than not, the most meaningful results will be in the form of descriptions of successful problem solving methods. Popular magazines, technical reports, and research resumés are just as likely to be retrieved—and desired—as the so-called "scholarly literature."

"Meatball searching" is the only way I can describe what I do. Most online searches will be executed in several databases, usually on more than one online service. With over 20 online services offering more than 500 databases at the disposal of the office, the cost of acquiring and maintaining thesauri for each becomes an expensive proposition, forgetting for the time being that many databases lack controlled vocabularies. Of course, the interactive nature of the online search is its biggest selling point, because you can construct the basic strategy, execute it, look at some preliminary results, modify the strategy, look at the "new and improved" results, and so on.

The most important question to ask the client during the interview process (well, perhaps the second most important question, after, "Who's going to pay for this?") is not "What?" but "Why?" By this I mean that if the search topic can be defined as a problem statement, the meatball searcher (or any other kind of searcher, for that matter) is quite likely to stumble upon a "hidden" aspect of the search.

For example, suppose a patron arrived (with hundred dollar bills dangling out of his or her pockets) and said, "I want a search to see if there's any connection between chemical x and lung cancer." A searcher with access to MEDLARS would probably start by properly identifying the specific chemical as well as the exact type of cancer, in order to be able to take full advantage of MEDLINE, the Toxicology Data Bank, TOXLINE, RTECS (the Registry of Toxic Effects of Chemical Substances), and so on. Searchers without MEDLARS might still go through this "refinement" process, anticipating a search on MEDLINE, BIOSIS, perhaps Chemical Abstracts, or whatever.

However, the information this patron needs may not necessarily be what he or she is asking for. That is why it is critical to ask, "Why?" If the patron is studying the toxic effects of chemical x, or researching the causes of lung cancer, fine. but if the patron is an attorney involved in a workers compensation suit, he or she may need to obtain other literature to find potential expert witnesses, legal precedents, and news accounts. On the other hand, if the patron is responsible for safety in a factory in which workers are exposed to this substance, what may really be required is information about correcting the source of the exposure, specifications for safety equipment, and so on.

Thus, the specific problem for which the information is sought is often just as important as the search subject itself.

Thus, the specific problem for which the information is sought is often just as important as the search subject itself.

Thus, the words and phrases used by the client are likely to constitute the first statements of the online search; the search subject, as well as the reason behind its request, will determine which databases are likely to yield useful information. Controlled vocabulary is used to retrieve broader materials later in the search process. Each database searched is therefore likely to result in at least two sets of offline prints—those obtained with the free-text search statements, and those additional citations retrieved by controlled vocabulary. When the online service features the option, the first part of the strategy (containing the free-text commands) is stored for execution against other databases. Thus, this reliance on free text searching might be regarded as the first point of "meatball searching." This procedure of obtaining a few fairly relevant citations and building upon this initial set by retrieving more items with alternate (i.e., controlled vocabulary) search terms has been referred to as "pearl culturing."

This strategy permits the creation of several sets, the first presumably the most relevant while subsequent sets may reflect somewhat less specific but nevertheless useful information.

Because of the inter-relationships of most technical subject areas, databases are selected according to both scope and content. The problem solving information required by clients of NASA/UK TAP's services is likely to have appeared in a variety of formats—dissertations government reports, the popular press, and so on—and thus NTIS and Magazine Index receive a fair amount of usage. Very few searches are conducted on only one database, and these tend to be nonbibliographic searches on databases like the Electronic Yellow Pages and other files with straight marketing data. This multi-database approach is often helpful in assuring that potentially relevant materials are not missed in "meatball searching." Moreover, such a search is likely to reflect the literature of several disciplines as they relate to the specific topic at hand.

Because many of the individuals representing client organizations are not necessarily technical experts—and, more importantly, because they are seeking problem-solving information—an unusual mix of databases is often employed. For example, a local agency recently requested information on employer-sponsored health awareness programs for employees. Among the databases searched were NTIS (because other local agencies may have developed such programs with Federal assistance), MAGAZINE INDEX (because innovations of local governments are more likely to be reported in the popular press long before an academic decides to write a "scholarly paper" on the event), ABI/INFORM (because the business community is usually interested in the pragmatic aspects of such programs), HEALTH PLANNING & ADMINISTRATION (for obvious reasons, although citations about hospitals' programs for employees were eliminated because of the difference in the kinds of services that were feasible); and NASA/RECON (because NASA's space flight centers, applications centers, and contractors are likely to have incorporated such programs, or may have assisted their clients in doing so). Obviously many other databases suggest themselves; the question is, how many files should be used?

Forget "comprehensive" and "complete." Most online databases only go back a few years; very few single-discipline databases cover absolutely every possible subject area or relevant publication; and regardless of how "good" you are, you have to accept the chance that something in the database is going to be missed. Are you willing to scan every title in a database of 300,000 items? I rather doubt it.

The "meatball" search, for both reasons outlined above, is likely to result in a large number of citations, emphasizing recall over precision (few of my searches provide less than 100 citations). These may provide access to information addressing social aspects of technical problems, or vice versa. This is to suggest that the meatball searcher is not likely to have a great deal of faith in: (a) the single-database search (except for nonbibliographic or directory searches, or for "quickie" searches used just to get a feel for how much there might be "in there"), because every search has (or should have, if it's been adequately defined) at least two aspects; or (b) the author search, because authors may write about a variety of related subjects from different angles and in different publications, some of which may not be indexed in what appears to be the "appropriate" database.

INFORMATION NEEDS

What do the clients of such a service need? Management and marketing information for new business ventures is always a hot topic. Among the recent searches I have conducted are these gems:

- use of earthworms as meat filler in beef (I would be *very* careful about where—and what—I eat when dining out after doing this search).
- marketing snake meat as a delicacy (see above parenthetical remark
- business outlook for gourmet popcorn (I bet you didn't know that Kentucky is among the leading producers of corn for popping).
- marketing of diaper services, pet sitting services, and "coal art" (carvings made from coal. . .well, it's high sulfur: you don't want us to burn it, do you?)

Other interesting subjects have included:

- clothing choices of the elderly
- the psychology of color in interior design, especially for churches and other places of worship
- mortuary and burial practices in the 18th and 19th Centuries.

Each of these, and almost all the other searches performed each year (800-1,000 annually), are for problem-solving information, as opposed to a client's idle curiosity or desire to write a "scholarly paper."

Part of the need for "meatball searching" is based on the clients and their problems. Unlike the setting of a hospital library—where patrons are looking for a few highly relevant citations and are, more importantly, well acquainted with online searching principles—NASA/UK TAP clients in many cases may not be subject specialists with access to the latest literature, and (in the experience of the author) seem even less likely to fully understand online information retrieval.

"Meatball searching" from my viewpoint seems more likely to occur in search centers serving many different disciplines and levels of client information sophistication. The generalist searcher, I would bet, is the obvious culprit.

LAWS OF MEATBALL SEARCHING & THE ADVERSARIAL APPROACH

Recently, while comparing notes with other multidisciplinary/multidatabase searchers as

►

to the nature of meatball searching, a specific characteristic was pointed out to me by Julia Redford of the U.S. Department of Energy's Office for Scientific and Technical Information (formerly the Technical Information Center) in Oak Ridge, Tennessee. Ms. Redford called it the "adversarial approach," and described its manifested attitude as it might be expressed during a typical search when things are not progressing smoothly:

"Damn it, I KNOW you're in there!" Implicit in this attitude is a belief that online databases and the search services on which they are searched are not to be regarded as close personal friends but as temporary allies which might turn on the searcher at any moment.

"Damn it, I KNOW you're in there!"

Implicit in this attitude is a belief that online databases and the search services on which they are searched are not to be regarded as close personal friends but as temporary allies which might turn on the searcher at any moment. Bad typing, poor choice of search terms, awkward strategies, typographical errors in the database, or any of about 817 other factors may provide zero postings during a search. Part of the problem may be due to the fact that the most sophisticated information retrieval services are not very user friendly (on the other hand, it can be suggested that what we really need are more friendly users).

Furthermore, I try to be careful in developing my search strategies, from selecting the databases to inputting the commands. When the results are not satisfying, I feel cheated, and my response is not disappointment or fear but frustration and anger:

• "Damn it, I KNOW you're in there!"

The searcher must always be prepared to call up any trick in the book to outsmart the database. Fight back. Don't be gentle. Take it personally.

Following are some general laws which may be applicable to recognizing the meatball searcher.

Jack's First Law: No matter how many databases you use, you will eventually settle into concentrating on the extensive use of a very few, using the others only as the specific obvious need arises. For me, these are NASA/RECON's File Collection D, NTIS (to pick up any other relevant government reports NASA has excluded), and MAGAZINE INDEX, because sooner or later somebody wrote an article that was published in *U. S. News & World Report*, or *Playboy*, or *National Geographic*. These might be regarded as a representative "core collection" of databases for the typical meatball searcher. Other databases are consulted according to the specific information need.

Maintaining the database documentation alone can be a full-time job. I edit a column on new and forthcoming databases for the Central/Eastern Kentucky Online User Group Newsletter. I will often receive some advance notice of new databases or changes to current files; I also try to follow the "Big Three" online publications, and look at the news messages on the online services when I sign on. These and other sources alert me to the industry's news.

Jack's Second Law: If your bookcases are not overflowing with user manuals and aids, your documentation is weak. If you have all the documentation for all the databases you use, you are not using enough databases. If you have access to all the databases you need, you are not a "meatball searcher."

No matter what kind of environment in which you operate, no matter what level of "information sophistication" your clients possess, and no matter what your subject area is and how narrowly you can define it, there will always be (a) some database, somewhere, that's just perfect for your need, that you don't access, and/or (b) some data or information that is simply not included in the databases you access, or in any database at all.

"Meatball searching" is the original "hacking." Remember, remote terminals have been around a lot longer than "affordable microcomputers." (Speaking of which, I once felt quite seriously that if you can afford a microcomputer and are smart enough to use it, you probably don't really need it. Now that I have been using a microcomputer at home for writing, and am expecting one to use in the office, my position has changed greatly).

"Meatball searching" presupposes clients with information needs that are not addressed by traditional information resources. Meatball searching with both sides of the brain, combining knowledge, luck, planning, intuitiveness, and what Judith Wanger has called "it." If you doubt you have "it," you're probably right.

Jack's Third Law: Anybody who can type can perform online searching.

Jack's Corollary to Jack's Third Law: Anybody who can hold a knife can perform brain surgery.

Any takers?

I type at about 45 to 50 words per minute, using a peculiar hunt-and-peck method that relies on striking the appropriate key at any moment with whatever part of my anatomy is not otherwise occupied with cigarettes, coffee, pencils, paper clips, and other office items. The advent of 1200 (and higher) baud rates sort of de-emphasizes the concern with human keyboard entry. How well one thinks is of much greater consequence than how quickly one types. Meatball searchers are not typists, although they may eventually learn to type quickly.

Just when you are ready to give up on a search, convinced that there is nothing relevant to the topic at hand, you will find exactly the item you are looking for. . .sometimes. Sometimes not. The "meatball searcher" accepts this outcome very reluctantly, and is likely to try to redefine the strategy and begin again. Frankly, I take the approach that none of our clients are all *that* original. Surely SOMEONE has thought of using earthworms as a meat extender. One must have a sense of humor—and a strong stomach—to deal with the ambiguities of this job.

Jack's Fourth Law: Ignore false hits. They are the speed bumps on the highway of life.

I recall not long ago searching for productivity improvement techniques for local government agencies. Naturally, LOGIN (Local Government Information Network) was consulted [7]. At the time, it was still called DEVELOP, and the Urban Technology Exchange was just one part of a file that included information on energy, appropriate technology, agriculture, and other subjects. One of the agriculture hits was retrieved by the search strategy—it referred to improving the productivity of bulls at stud. Specifically, it provided a rather graphic explanation of the technique for measuring the reproductive organs of the bull, and included a warning about exercising "great care." I would imagine that during such a procedure this would be excellent advice.

Jack's Fifth Law: Practice makes perfect. Nobody practices. Therefore, nobody's perfect.

There are going to be times that, for one reason or another, a good item is missed in any search. These things happen. I personally work on the assumption that the materials I do retrieve will be at least as good as whatever I miss. Confidence in yourself becomes something of a self-fulfilling prophecy. Regardless of the search topic or the databases you use, you are not ever going to get *everything*. Most databases only cover a few years of literature, and there are a lot of materials out there that, for one reason or another, will probably never show up in a database. Thesaurus terms might be too narrow or too broad, and the free-text terms selected may not match the vocabulary of the author or the abstracter.

Nobody is asking you to be perfect—merely adequate. Strive for "the perfect search" each time you sign on, given limits of time, money, and other resources. I believe that self-confidence becomes a self-fulfilling prophecy, because it inspires the searcher to calculate the risks and take educated chances and thus gain experience.

MISCELLANEOUS ATTITUDES & CHARACTERISTICS

I am not certain that I have met very many other meatball searchers (although I have run into a few searchers who are meatballs—you know who you are). I have been told that I am arrogant, loud, rude, vulgar, and generally obnoxious—but I have my bad points, too. After all, nobody's perfect.

The meatball searcher, among other things, really enjoys online searching: he or she enjoys the challenge of identifying the topic at hand, exploring new databases, providing answers. The search assumes the appearance of a well-played game of chess, at least as long as the process is proceeding smoothly. On the other hand, when searcher and database assume the adversarial relationship, the competition becomes the intellectual equivalent of a street brawl, often including (in my case, at least) swearing at the database, the online service, the telecommunications network, the terminal, the modem, the indexing/abstracting staff, and anyone else who may be vaguely perceived as responsible for undesirable search outcomes. When offline, the meatball searcher spends a great deal of time in closely related activities, such as preparing management information reports, verifying search invoices, maintaining documentation, readying the information retrieval literature, preparing for searches, participating in user group activities, and dreaming of new and interesting database possibilities.

The meatball searcher is tenacious but knows when to quit. Even so, no search request is ever really closed, especially when the retrieved results have been disappointing. When new databases are announced as available or forthcoming, do you suddenly remember previous searches which would have been "perfect" for the new file? At the same time, are you skeptical about such product announcements, especially when the producer suggests that its

►

file is the one and only database you'll ever need?

Jack's Sixth Law: Healthy skepticism can be a positive attitude.

How the searcher describes him/herself, I think, has a lot to do with whether that person can be properly called a meatball searcher. A number of my personal acquaintances seem to think of themselves as librarians who "do" online searching from time to time. I am online at least every day, and often for several hours at a time. Since online information retrieval has been traditionally associated with libraries, it is natural to expect that many practitioners would use that particular title. On the other hand, you could sit in a garage for twenty years, but that will still not make you either a mechanic or an automobile.

My own case may be an excellent example. Sure, I have a Master's degree in Library and Information Service. (Incidentally, this degree is not accredited by the American Library Association, but I don't mind; if the ALA doesn't try to tell me how to do online searching, I won't try to tell them how to do whatever it is they're supposed to do). I also have a Bachelor's degree in International Relations; that doesn't make me a diplomat.

CONCLUSIONS

Donna Dolan and Michael Kremin listed some of the important characteristics of good searchers. These traits and skills are:

- concept analysis
- flexibility of thinking
- ability to think in synonyms
- anticipation of variant word forms and spellings
- self-confidence [8].

I would expect that a searcher possessing these traits, and who knows when to take the adversarial approach, is a meatball searcher. This is just a hunch, but I wonder if some of the authors of works I've cited would confess, "Yes, I'm unorthodox! Yes, I'm nontraditional! Yes, I'm a meatball searcher!"

Psychologically, are meatball searchers somehow "different?" I'm not sure. I recently got a reprint request for my LOGIN article *(Database,* June 1984) from a psychiatric research center. That should tell me something, I suppose. It is quite possible I've been online too long.

NOTES

[1] Oldroyd, Betty K., and Citroen, Charles L. "Study of Strategies Used in On-Line Searching." *ONLINE REVIEW*, 1(4), December 1977, pp. 295-310.

[2] Fenichel, Carol Hansen. "An Examination of the Relationship Between Searching Behavior and Searcher Background." *ONLINE REVIEW*, 4(4), December 1980, pp. 341-347.

[3] Howard, Helen. "Measures that Discriminate Among Online Searchers with Different Training and Experience." *ONLINE REVIEW*, 6(4), August 1982, pp. 315-327.

[4] Rholes, Julia M., and Droesser, Judith B. "Online Database Searchers: Cognitive Style." *NATIONAL ONLINE MEETING PROCEEDINGS*—1984. Medford, N. J.: Learned Information, Inc., 1984. pp. 305-311.

[5] Dolan, Donna R. "Conceptualizing Online Search Requests."*(Database*, 6(1), February 1983, pp. 77-78.

[6] Wanger, Judith. "Multiple Database Use: The Challenge of the Database Selection Process." *ONLINE*, 1(4), October 1977, pp. 35-39.

[7] Jack, Robert F. "Searching LOGIN, the Local Government Information Network." *DATABASE*, 7(2), June 1984, pp. 35-43.

[8] Dolan, Donna R., and Kremin, Michael C. "The Quality Control of Search Analysts." *ONLINE*, 3(2), April 1979, pp. 8-16.

THE AUTHOR

ROBERT F. JACK is the Technical Information Services Coordinator at the NASA/University of Kentucky Technology Applications Program (NASA/UK TAP). He has a Master of Arts in Library and Information Services (1978, University of Toledo, Community Information Specialist Program). His previous contributions have included articles in both *Online and Database,* conference presentations, and other creative writings such as grocery lists, income tax returns, and letters to Mom.

Communications to the author should be addressed to Robert F. Jack, Technical Information Services Coordinator, NASA Scientific & Technical Information Facility, P.O. Box 8757, Baltimore/Washington International Airport, MD 21240, 301/859-5300 ext. 107

THE Friendly User

by Lucinda D. Conger
U.S. Dept. of State Library

GATEWAYS: AN INTRODUCTION*

HOST SYSTEMS AND GUEST SYSTEMS

We've been hearing a lot about gateways these days. OCLC has one with BRS; RLIN has one with DIALOG; MDC (Mead Data Central) has one with DIALOG; DataTimes has one with Dow Jones; WESTLAW has one with VU/TEXT. A gateway is more than a single telephone call to a favorite and familiar online service. It is a matter of access: telecommunications access. It is like that favorite sales pitch: Buy one, get one free. You arrange for access to OCLC, RLIN, MDC, DataTimes, or WESTLAW, and *voilá*, you have access to the other online service. For convenience, I will call the system offering the gateway the host system, and the system available through the gateway, the guest system. Because, you see, most gateway arrangements are one-way. True, BRS offers an OCLC database, but that is not a dial-through service to the OCLC cataloging utility.

With a gateway arrangement, you do not have to set up additional accounts with BRS, DIALOG, Dow Jones, or VU/TEXT. The host service manages the accounting. And you only need to keep track of one password, one set of logon protocols. Unlike the sales pitch, however, you do not really get the guest service for free. You do have to pay for the databases you use, but you pay through the host system, and often the host system will provide telecommunications at its own rate, not at the rate charged if you accessed the guest search service directly.

In the simplest and most direct kind of gateway, you, the user, also have to learn the search language of the guest system into which you are going. Often, the languages of the host system and the guest system are different. This is true of the gateway between MDC and DIALOG, between RLIN and DIALOG, and may be true of the relationship between DataTimes and Dow Jones, although that depends on whether you are talking to a Dow Jones sales rep or the DataTimes sales rep. Dow Jones says that you will be searching with its own languages, DataTimes says you can search Dow Jones with its search language. Maybe, the choice of language will be an option for you, the searcher.

INTELLIGENT GATEWAYS

Some host services do go one step farther. Instead of merely providing telecommunications access to another system, they will help you use it. The help is usually provided in the form of a menu or help screens, which walk you through the search process. This kind of gateway is called an intelligent gateway. The OCLC gateway to BRS is an intelligent gateway.

EASYNET

So far, I have been talking about familiar host systems: hosts whose use have intrinsic value, aside from the access they can provide to other online systems. But there are online systems that exist simply as gateways, the Grand Central Stations of online searching. Of these, the EasyNet family of gateway products is the most well-known. Western Union's InfoMaster and CompuServe's IQUEST are both variations on the same gateway product, Telebase's EasyNet. EasyNet is the arranger of relationships *par excellence*. It and its abovementioned variants do all the paperwork and provide the telecommunications access. All you need is one telephone number and one password to EasyNet, and EasyNet does the rest. It keeps track of what guest services you use, it sends you the bill, and you pay it the money owed.

As for searching, EasyNet provides two modes of operation. Mode I is strictly for beginners; it is a very intelligent gateway that not only chooses the database, but also chooses the system on which the search is to be run. True, most of the searches appear to be run on DIALOG and BRS, but EasyNet now boasts access to 15 guest systems, with more than 700 databases. In mode II, the searcher can select the desired guest system.

The only problem with EasyNet, to my way of thinking, is the way it is priced. For many people unfamiliar with any kind of online searching, the pricing system may seem fair: you only pay for records retrieved. You do not pay for searches which result in no hits; you do not pay by the time you are online (at least not in terms of database charges, although you do in terms of communications charges). However, if you retrieve more than ten records, you pay a tariff for viewing the extra records, and you have to pay extra to display the abstracts of even the first ten items retrieved. Since the number of records retrieved is no more predictable than the time you might be online, predictability of costs should not be a deciding factor in using this service.

---

*Presented at a seminar on gateways sponsored by FEDLINK held at the Library of Congress, February 20, 1987.

DATABASE CHOICE ON EASYNET

Rather its major value lies in making the database choice for the inexperienced user. This choice is, I think, the biggest problem with online searching. New online searchers of any breed, professional and non-professional alike, have this problem. Given a set of questions, what is or are the most appropriate databases in which to search for the answer?

It is my belief that any searcher who has been able to figure out how to get online in the first place using a piece of telecommunications software can usually figure out how to use an online searching system. Aside from such search languages as those used by **DROLS** or **NASA/RECON**, most search languages are not so arcane that they cannot be mastered to the extent that simple searches can be run. But how to decide in which database to run them?

Most online systems do provide some database selection tools as part of their system. This is the function of such files as **CROS** on BRS and **DIALINDEX** on DIALOG. Some software packages, such as Pro-Search, have also tried to provide help in this area. EasyNet, using the techniques of artificial intelligence, couched in menus and system queries, actually makes the database choice for the user. Whether the choice is the right one or the most appropriate one is another question. If the searcher gets some results from the search with which he or she is happy, then the choice was correct. If not, that is not to say the selection was wrong; it could simply be the classic reference problem of the searcher not asking for what he really wanted.

SUBSCRIPTION GATEWAYS

At the risk of confusing you, I do want to mention the subscription gateway. This is not an actual gateway, as I have just described, based on a telecommunications link. Instead, it is a paper link based on the willingness of a host service to do all the paperwork and make all the accounting arrangements, so that you can get access to another system. The host system provides you with the telephone number, the logon protocols, the password, and the system documentation, and bills you for the guest system's use as part of its own billing to you.

This kind of arrangement is particularly valuable when an American search system agrees to play host to a European search system. Although it would be nice to be able to dial through an American system to a European system, because European access telephone numbers are usually 12 digits long, it is not really necessary if you have communications software in which you can store such a telephone number. This is the kind of gateway VU/TEXT provides for Datasolve. In this case, VU/TEXT is acting like a real estate agent, promoting and providing access to properties which it does not actually own.

TRANSPARENT GATEWAYS

There is also another kind of gateway, a transparent gateway, which refers to services offered within the same host system, but which happen to be on different computers. This kind of gateway, being internal to the host system, is not visible or obvious to the user, and is therefore referred to in the literature as a "transparent" gateway. The **Official Airline Guide**, for instance, is available via this kind of gateway to users of DIALOG, Dow Jones, and all the other hosts that offer it. Dow Jones text files are also linked to its numeric files via a gateway, visible only because of the difference in search techniques. In fact, differences in search technique may be the only clue to the searcher that a transparent gateway is in use.

As we have seen, gateways provide a convenience to users of any system that is willing to play host to any other guest system or service. The host system is not only making sophisticated use of telecommunications technology, but is also expanding the range of the services it can offer to its users. Benefits accrue, thereby, to both the host and the guest. Some potential customers of EasyNet have expressed concern for the viability of such guest services as DIALOG as independent search services. Never fear! The current joke is that EasyNet's DIALOG bill is so big, an entire UPS truck is required to deliver it.

In fact, some smaller guest services with less developed marketing strategies may benefit considerably from being included in a gateway arrangement. However, adequate documentation is the key to their use. Both host and guest services should work together to provide scope notes for each file, and customers should alert the host service, in particular, to the information they need to make full use of the guest services offered.

Communications to the author should be addressed to Lucinda D. Conger, U.S. Department of State Library, Room 3239NS, Washington, DC 20520; 202/632-0486.

ONLINE AND OFFLINE PRINTING . . . RELATIVE COSTS

by Brian Williams
Microcomputer Coordinator
Reference Department, Kennedy Library
California Polytechnic State University

There are several factors in choosing whether to use a microcomputer for searching or not. One of the advantages is getting the retrieval in machine readable form for later processing. One of the disadvantages is that while you are printing online, or downloading, you are paying the regular connect hour prices, as opposed to getting this free if you request offline prints.

This study is a projection of the differences in costs for online printing and offline printing. Two typical searches are represented for this comparison. One is a 15 minute search with 10 citations and the other is a 15 minute search with 30 citations. Online retrieval is compared to offline retrieval in these two situations. I timed the downloading process using an IBM PC with Smartcom II software and a Decwriter printing terminal in the equivalent of DIALOG's format 3. Each of these machines is equipped with a Hayes 1200 baud modem. In several tests of both BRS and DIALOG, the time for online printing varied from 2 to 3 seconds per citation on either machine. For the purposes of this study, I used 3 seconds per citation in the calculations of the time required for online printing. This seems reasonably conservative, but can only be used for comparison purposes. The actual time required for downloading depends on several factors including the load on the system, differences in record size, communications software, and the like. With this in mind, the projections of costs given in this article are not indications of what an actual search will cost *exactly*, but a comparison of search costs for online and offline printing in the two systems based on a consistent set of assumptions.

The costs for online time per database and offline and online printing are taken from June 1986 databank documentation. The offline print charges that BRS quotes do not include the per page charge of $.23 or the postage that is included in the final bill. As DIALOG does not explicitly have similar charges, we have added $.06 to the per citation charge for the source field that BRS quotes in their database **FILE** to include these costs.

The online rate used for the BRS charges is the base rate of $16 per hour, corresponding to the Sub Level 4 rate on their documentation, which is what we pay here at the Kennedy Library. Similarly, the DIALOG database online rates have been adjusted by deducting $5.00, which represents our discount (approximately). If you do not enjoy such a discount, then the DIALOG costs would be closer to real ones if you add a dollar per search, and the BRS costs would be much higher.

$9.00 per hour is added into the formula for the search cost to account for communications charges.

The two tables in this report have been generated from an Appleworks database to compare the costs of offline retrieval to online retrieval. Both DIALOG and BRS files are included, as they have different files and different costs.

In Table 1, column A is the database name, B is the databank, C is the databank's code name or number for that database, D is the adjusted cost of accessing the database per hour, E is the cost of a 15 minute search without any retrieval, F is the offline charge per citation, G is the total charge of a 15 minute search plus the cost of having 10 citations printed offline and mailed, and H is the same cost for 30 citations.

TABLE 1

OFFLINE RETRIEVAL COSTS

A	B	C	D	E	F	G	H
Database name	Databank	Code	Cost/ Hr.	Cost/ Search	$/off Cit.	$/10 off.	$/30 off.
ABI/INFORM	BRS	INFO	51	15.00	.41	19.10	27.30
ABI/INFORM	DIALOG	15	76	21.25	.60	27.25	39.25
AGRICOLA	BRS	CAIN	20	7.25	.08	8.05	9.65
AGRICOLA	DIALOG	10+	34	10.75	.20	12.75	16.75
AGRIS INT'L	DIALOG	203	40	12.25	.15	13.75	16.75
BIOSIS	DIALOG	5+	79	22.00	.34	25.40	32.20
BIOSIS PREVIEW	BRS	BIOL	56	16.25	.18	18.05	21.65
BUSINESS SOFTW	DIALOG	256	85	23.50	.90	32.50	50.50
CA SEARCH	BRS	CHEM	70	19.75	.31	22.85	29.05
CA SEARCH	DIALOG	311+	85	23.50	.35	27.00	34.00
CAB ABSTRACTS	DIALOG	50	50	14.75	.35	18.25	25.25
CHEMNAME	DIALOG	301	153	40.50	.32	43.70	50.10
COMPENDEX	BRS	COMP	71	20.00	.33	23.30	29.90
COMPENDEX	DIALOG	8	105	28.50	.47	33.20	42.60
COMPUTER DATAB	DIALOG	275	91	25.00	.55	30.50	41.50
DISS ABS ONLIN	DIALOG	35	67	19.00	.25	21.50	26.50
DOE ENERGY	DIALOG	103+	73	20.50	.25	23.00	28.00
EMBASE	BRS	EMED	57	16.50	.20	18.50	22.50
EMBASE	DIALOG	72+	79	22.00	.33	25.30	31.90
ENERGYLINE	DIALOG	69	103	28.00	.35	31.50	38.50
ENVIROLINE	DIALOG	40	103	28.00	.35	31.50	38.50
ENVIRONMTL BIB	DIALOG	68	55	16.00	.15	17.50	20.50
ERIC	BRS	ERIC	16	6.25	.06	6.85	8.05
ERIC	DIALOG	1	25	8.50	.14	9.90	12.70
EXCEPTIONAL CH	BRS	ECER	26	8.75	.11	9.85	12.05
EYE-RTL	DIALOG	504	55	16.00	.20	18.00	22.00
FAMILY RESOURC	BRS	NCFR	41	12.50	.21	14.60	18.80
FAMILY RESOURC	DIALOG	291	52	15.25	.25	17.75	22.75
HEALTH PLAN &	DIALOG	151	31	10.00	.20	12.00	16.00
HEALTH PLANNING	BRS	HLTH	20	7.25	.07	7.95	9.35
INSPEC	BRS	INSP	62	17.75	.31	20.85	27.05
INSPEC	DIALOG	12+	91	25.00	.46	29.60	38.80
ISMEC	DIALOG	14	79	22.00	.40	26.00	34.00
LEGAL RES	DIALOG	150	85	23.50	.20	25.50	29.50
LEGAL RES	BRS	LAWS	83.5	23.12	.31	26.23	32.42
LISA	DIALOG	61	70	19.75	.25	22.25	27.25
MAGAZINE ASAP	BRS	MSAP	79	22.00		22.00	22.00
MAGAZINE INDEX	BRS	MAGS	79	22.00		22.00	22.00
MANAGEMENT CON	DIALOG	75	85	23.50	.55	29.00	40.00
MANAGEMENT CON	BRS	MGMT	58	16.75	.36	20.35	27.55
MATHSCI	BRS	MATH	46	13.75	.25	16.25	21.25
MATHSCI	DIALOG	239	67	19.00	.46	23.60	32.80
MEDLINE	BRS	MESZ	20	7.25	.07	7.95	9.35
MEDLINE	DIALOG	152+	31	10.00	.20	12.00	16.00
MENU-INTL SOFT	DIALOG	232	55	16.00	.15	17.50	20.50
MICROCOMPUT IN	DIALOG	233	40	12.25	.15	13.75	16.75
MLA BIBLIOGRAP	DIALOG	71	50	14.75	.15	16.25	19.25
NATIONAL NEWSP	BRS	NOOZ	79	22.00	.31	25.10	31.30
NATL NEWSPAPER	DIALOG	111	79	22.00	.20	24.00	28.00
NTIS	DIALOG	6	64	18.25	.25	20.75	25.75

►

TABLE 1 (continued)

NURS & ALLD HL	DIALOG	218	49	14.50	.25	17.00	22.00
OCEANIC ABSTRA	DIALOG	28	82	22.75	.45	27.25	36.25
PAIS INTER.	BRS	PAIS	46	13.75	.21	15.85	20.05
PAIS INTER.	DIALOG	49	64	18.25	.30	21.25	27.25
POLLUTION ABS	DIALOG	41	79	22.00	.40	26.00	34.00
POLLUTION ABS	BRS	POL	51	15.00	.26	17.60	22.80
PSYCINFO	DIALOG	11	50	14.75	.20	16.75	20.75
PSYCINFO	BRS	PSYC	31	10.00	.10	11.00	13.00
PTS F&S INDEX	BRS	PTSI	81	22.50	.21	24.60	28.80
PTS PROMT	DIALOG	16	109	29.50	.58	35.30	46.90
ROBOTICS INFOR	BRS	RBOT	56	16.25	.29	19.15	24.95
SCISRCH (sub)	DIALOG	34+	58	16.75	.25	19.25	24.25
SOC SCISRCH	DIALOG	7	64	18.25	.35	21.75	28.75
SOC SCISRCH	BRS	SSCI	51	15.00	.16	16.60	19.80
SOCIOLOGICAL A	DIALOG	37	55	16.00	.30	19.00	25.00
SOCIOLOGICAL A	BRS	SOCA	41	12.50	.21	14.60	18.80
SOVIET SCIENCE	DIALOG	270	60	17.25	.50	22.25	32.25
SPORT DATABASE	BRS	SFDB	46	13.75	.16	15.35	18.55
ZOOLOGICAL REC	DIALOG	185	82	22.75	.26	25.35	30.55

TABLE 2

ONLINE RETRIEVAL COSTS

In Table 2, columns A, B, C, D, and E are the same as in Table 1. Column F is the online display charge per citation, G is the total search cost for a 15 minute search and 10 citations printed online, and H is the same thing for 30 citations.

A	B	C	D	E	F	G	H
Database name	**Databank**	**Code**	**Cost/ Hr.**	**Cost/ Search**	**$/On. Cit.**	**$/10 On.**	**$/30 On.**
ABI/INFORM	BRS	INFO	51	15.00	.39	19.40	28.20
ABI/INFORM	DIALOG	15	76	21.25	.50	26.96	38.38
AGRICOLA	BRS	CAIN	20	7.25	.09	8.39	10.68
AGRICOLA	DIALOG	10+	34	10.75	.10	12.11	14.83
AGRIS INT'L	DIALOG	203	40	12.25		12.66	13.48
BIOSIS	DIALOG	5+	79	22.00	.24	25.13	31.40
BIOSIS PREVIEW	BRS	BIOL	56	16.25	.22	18.99	24.48
BUSINESS SOFTW	DIALOG	256	85	23.50	.90	33.28	52.85
CA SEARCH	BRS	CHEM	70	19.75	.29	23.31	30.42
CA SEARCH	DIALOG	311+	85	23.50	.23	26.58	32.75
CAB ABSTRACTS	DIALOG	50	50	14.75	.25	17.74	23.73
CHEMNAME	DIALOG	301	153	40.50	.18	43.65	49.95
COMPENDEX	BRS	COMP	71	20.00	.36	24.27	32.80
COMPENDEX	DIALOG	8	105	28.50	.35	32.95	41.85
COMPUTER DATAB	DIALOG	275	91	25.00	.50	30.83	42.50
DISS ABS ONLIN	DIALOG	35	67	19.00	.25	22.13	28.40
DOE ENERGY	DIALOG	103+	73	20.50	.25	23.68	30.05
EMBASE	BRS	EMED	57	16.50	.42	21.25	30.75
EMBASE	DIALOG	72+	79	22.00	.44	27.13	37.40
ENERGYLINE	DIALOG	69	103	28.00	.30	31.93	39.80
ENVIROLINE	DIALOG	40	103	28.00	.30	31.93	39.80
ENVIRONMTL BIB	DIALOG	68	55	16.00		16.53	17.60
ERIC	BRS	ERIC	16	6.25	.11	7.56	10.17
ERIC	DIALOG	1	25	8.50	.10	9.78	12.35
EXCEPTIONAL CH	BRS	ECER	26	8.75	.07	9.74	11.73
EYE-RTL	DIALOG	504	55	16.00	.20	18.53	23.60

TABLE 2 (continued)

FAMILY RESOURC	BRS	NCFR	41	12.50	.27	15.62	21.85
FAMILY RESOURC	DIALOG	291	52	15.25	.25	18.26	24.28
HEALTH PLAN &	DIALOG	151	31	10.00	.05	10.83	12.50
HEALTH PLANNIN	BRS	HLTH	20	7.25	.07	8.19	10.07
INSPEC	BRS	INSP	62	17.75	.20	20.34	25.52
INSPEC	DIALOG	12+	91	25.00	.36	29.43	38.30
ISMEC	DIALOG	14	79	22.00	.30	25.73	33.20
LEGAL RES	DIALOG	150	85	23.50	.10	25.28	28.85
LEGAL RES	BRS	LAWS	83.5	23.12	.20	25.90	31.44
LISA	DIALOG	61	70	19.75	.25	22.91	29.23
MAGAZINE ASAP	BRS	MSAP	79	22.00	0.00	22.73	24.20
MAGAZINE INDEX	BRS	MAGS	79	22.00		22.73	24.20
MANAGEMENT CON	DIALOG	75	85	23.50	.48	29.08	40.25
MANAGEMENT CON	BRS	MGMT	58	16.75	.37	21.01	29.53
MATHSCI	BRS	MATH	46	13.75	.25	16.71	22.62
MATHSCI	DIALOG	239	67	19.00	.32	22.83	30.50
MEDLINE	BRS	MESZ	20	7.25	.07	8.19	10.07
MEDLINE	DIALOG	152+	31	10.00	.05	10.83	12.50
MENU-INTL SOFT	DIALOG	232	55	16.00		16.53	17.60
MICROCOMPUT IN	DIALOG	233	40	12.25	.15	14.16	17.98
MLA BIBLIOGRAP	DIALOG	71	50	14.75		15.24	16.23
NATIONAL NEWSP	BRS	NOOZ	79	22.00	.20	24.73	30.20
NATL NEWSPAPER	DIALOG	111	79	22.00	.10	23.73	27.20
NTIS	DIALOG	6	64	18.25	.25	21.36	27.58
NURS & ALLD HL	DIALOG	218	49	14.50		14.98	15.95
OCEANIC ABSTRA	DIALOG	28	82	22.75	.40	27.51	37.03
PAIS INTER.	BRS	PAIS	46	13.75	.22	16.41	21.72
PAIS INTER.	DIALOG	49	64	18.25	.20	20.86	26.08
POLLUTION ABS	DIALOG	41	79	22.00	.30	25.73	33.20
POLLUTION ABS	BRS	POL	51	15.00	.29	18.40	25.20
PSYCINFO	DIALOG	11	50	14.75	.35	18.74	26.73
PSYCINFO	BRS	PSYC	31	10.00	.22	12.53	17.60
PTS F&S INDEX	BRS	PTSI	81	22.50	.29	26.15	33.45
PTS PROMT	DIALOG	16	109	29.50	.48	35.28	46.85
ROBOTICS INFOR	BRS	RBOT	56	16.25	.28	19.59	26.28
SCISRCH (sub)	DIALOG	34+	58	16.75	.20	19.31	24.42
SOC SCISRCH	DIALOG	7	64	18.25	.30	21.86	29.08
SOC SCISRCH	BRS	SSCI	51	15.00	.18	17.30	21.90
SOCIOLOGICAL A	DIALOG	37	55	16.00	.20	18.53	23.60
SOCIOLOGICAL A	BRS	SOCA	41	12.50	.23	15.22	20.65
SOVIET SCIENCE	DIALOG	270	60	17.25	.40	21.83	30.98
SPORT DATABASE	BRS	SFDB	46	13.75	.15	15.71	19.62
ZOOLOGICAL REC	DIALOG	185	82	22.75	.20	25.51	31.03

CONCLUSIONS

Certainly, with the given assumptions, the differences between the costs of online and offline printing are not significant, with some exceptions. In general, we can say that for some sort of normal search, downloading is not disadvantaged by the fact that we have to pay relatively high connect charges for the process. Further, under the circumstances, the advantages of getting the retrieval immediately might make the online printing option preferable even without the added advantage of getting the retrieval in machine readable form.

Communications to the author should be addressed to Brian Williams, Kennedy Library, Cal Poly State University, San Luis Obispo, CA 93407.

SEARCHWORKS: DOES IT REALLY WORK?

by Bill Coons
Current Awareness Services Librarian
Albert R. Mann Library
Cornell University

INTRODUCTION

The perfect online search software, which does not exist, should have powerful, versatile and uncomplicated telecommunications, uploading, accounting, searching, and post-processing capabilities. Barbara Quint, writing in the January '86 issue of *ONLINE*, succinctly summarizes the present status of this type of software: "No package has yet come to market that does everything well." To wait for the perfect package, however, is to wait forever. There are various adequate online search aids presently available and more are being developed every year. They vary in their strengths and weaknesses; their suitability and usefulness depends upon the match between your particular needs and the software's specific capabilities. The bottom line is this: if the package does not save you time and money, streamline the searching process, improve the quality of the product and/or offer some different, innovative and useful applications, you do not want it. This review will attempt to provide you with enough information about one particular product, SearchWorks, for you to compare it to other online search aids and decide its applicability for your particular searching environment.

PRODUCT OVERVIEW

OnLine Research Systems, Inc., a small, privately-held company based in New York City, has recently released SearchWorks, a concise online searching aid designed to incorporate various aspects of post-search processing into the standard capabilities expected of front end software. Written for the IBM PC, XT and AT (or fully compatibles), SearchWorks performs best on a hard disk. It may be used to access any online system, but the applications it was designed for only work in bibliographic databases carried by BRS, DIALOG, SDC and NLM. Standard capabilities include automatic logon and other telecommunication functions, uploading, type-ahead, search statement recall and edit, scrolling, accounting, cover sheets, and basic online search assistance with system commands and syntax. Help with specific database use or selection, such as offered in Pro-Search, is not provided.

The package's most unique and useful feature is its post-search processing ability and the various sub-components which support it. Templates (a structured form you complete by entering replacement words for field labels) for tagged databases can be established and saved for future application to other searches. They translate traditionally inscrutable two-letter field tags into natural language, thereby providing clients with a comprehensible printout. It is also possible to clean up the search by deleting false drops, and the results can either be sorted (by any field tag you specify), or ranked (based upon a word=weight algorithim you establish). SearchWorks cannot, however, do both at the same time; it is an either/or situation. In addition, this package can generate an index to the search (useful for quickly locating references with primary keywords), and it allows for the merging of up to thirteen different database searches into one processed file.

These post-search processing capabilities, especially ranking, offer exciting new service challenges for the online searcher and a higher quality end product for the patron. Rather than just giving a patron an unconsolidated gaggle of references, SearchWorks makes it possible for the searcher to deliver an orderly list ranging from the most pertinent information to the least substantive, with irrelevant references relegated to the proverbial circular file.

DOCUMENTATION

With SearchWorks you receive a 180+ page manual and four quick-reference pages. The manual, written in a straightforward manner, but without graphics or adequate screen facsimiles, almost captures the essence of the program and is often enjoyably forthright ("To operate the computer when the disk is near capacity is to flirt with disaster.") It assumes the user knows how to choose vocabulary and formulate search strategies, and it contains a fairly complete and lucid orientation to the keyboard and use of the function and editing keys. The index is confusing because it refers to chapter sections and not individual page numbers in the manual. The quick-reference pages partially duplicate sections of the manual, but for the most part contain useful and important information not listed elsewhere. Since few of the screens and menu options are reproduced in the documentation you will have to print several screens for reference purposes. Questions which remain unanswered can be checked with the company. OnLine Research Systems staffs a support desk from Monday through Friday, 1:00 to 5:00 PM. EST, (212) 249-4308.

SETTING UP THE SOFTWARE

Initial users of SearchWorks do not need to go into a great level of detail to be able to begin to use the program. A two-hour investment of time (to read portions of the manual and complete the actual installation and configuration), and a healthy dose of the "I'll learn as I go" approach should get the program configured and installed. Instructions for both are clearly explained in the manual and the configuration itself is menu-driven. It leads you through thirty-four questions which set various parameters such as passwords, modem speed, and whether you are using floppy or hard disk. While there is no way to page back to review or correct a previous entry during configuration, once the initial profile is established you can quickly and easily modify it at any future time.

Once the configuration is completed, the installation program must be run. This establishes the software on your hard disk or makes working floppy copies of the program for PC users (with the floppy version, all of the start-up files cannot be placed on the disk, so you have to use the same DOS disk for start-up every time). The order of the configuration and installation seems somewhat backward because you have to configure the master disk (and chance the possibility of a fatal error) before you can backup (install) the master disk onto a floppy or the hard disk. Installation and use of the program on a hard disk is amazingly easy; floppy disk users have more restrictions due to file size limitations and the vulnerability of the disk. Because of space limitations, reduced flexibility, and the need to regularly change system disks, SearchWorks is recommended for use only on hard disk systems.

> **Installation and use of the program on a hard disk is amazingly easy; but floppy disk users have more restrictions due to file size limitations and the vulnerability of the disk. Because of space limitations, reduced flexibility, and the need to regularly change system disks. SearchWorks is recommended for use only on hard disk systems.**

ONLINE SEARCH GUIDE

The online search guide contains key search commands, syntax and examples for the supported online services. The search guide can be used both online and offline; a unique feature of the guide involves taking any one highlighted line from the search guide and transferring that information to the bottom of your search screen. The utility of this online guide is that it is quicker to search for these aids in this mini-database than through the various print sources scattered about a searching workstation. While it contains basic how-to references for online system specific commands, the online search guide is not the DIALOG Blue Pages or the BRS Aid Pages online. Unlike Pro-Search, SearchWorks does not help the searcher select or use specific databases.

The best way to get into the guide is to use the index and browse until you locate the specific area you wish to know more about. There are some bugs in this portion of the software and not quite as many access points as there should be. For example, the BRS 'root' command cannot be located on a search of the word 'root'; it can be located, however, by browsing through the index. ►

SEARCHWORKS FUNCTIONS

When SearchWorks is first loaded it presents you with a cryptic message concerning logon to one of the four online systems. If you do not choose to logon, a main menu with ten options appears (Figure 1). These options are edit search strategy (to prepare a search strategy offline for uploading), call (to logon to an online system), process search results (to reformat a search with templates, rank, sort or index the search, process cost and the cover sheet), file management (to copy, display, print, erase, list, or rename files), print a file, accounting and reports, record editor (to edit a downloaded file), modify, profile (to revise phone numbers, passwords and other defaults and profile information), help and exit. These functions and files will be discussed in the following sections: uploading and logon, downloading, file management, templates, rank/sort/index, and accounting and cover sheets.

UPLOADING AND LOGON

Logon with SearchWorks is quick and practically effortless; one key stroke is all that is needed to access any of the four supported online systems (in contrast, SmartCom II requires a minimum of three keystrokes). Passwords are masked and the user never sees the communication process. When SearchWorks is initially loaded, free disk space is checked and if less than 300K remains a warning is issued.

The initial screen asks for the letter of the online system to search. Entering **b,d,n** or **s** will result in an automatic logon to the appropriate online system. **0** will begin the logon procedure to online systems other than the supported ones. Because a logon profile for these other online systems has not been stored in your user profile, it is necessary to set the baud rate (1200 or 2400), duplex (half or full), parity (none, odd or even), data bits (7 or 8) and the phone number to dial. These parameters are not saved and must be re-entered each and every time you seek to search an online system other than those supported by SearchWorks. In contrast, SmartCom II has macros for the storage of up to 26 different online system phone numbers and passwords; Pro-Search has room for twenty-three.

A logon profile must be re-entered each and every time you seek to search an online system other than those supported by SearchWorks. In contrast, SmartCom II has macros for the storage of up to 26 different online system phone numbers and passwords; Pro-Search has room for twenty-three.

Uploaded search strategies can be saved, and if installed on a hard disk, there is no limit to the number of files which may be stored in SearchWorks' uploading files. The search strategy can be entered on up to 99 discrete lines, with a line limit of 76 characters per line (wrap-around is not possible; the Boolean 'or' must be used with set numbers to accommodate long search statements). Pro-Search's High Level interface allows 98 lines of 50 characters each (without wrap-around) and its Native Mode permits 50 lines of 240 characters each (with wrap-around). Also, in SearchWorks the upload must run as the first search statement; it cannot be entered at another position in the search. Special features include type-ahead and recall/edit of last line sent to the host.

DOWNLOADING

In SearchWorks you do not have to explicitly open a file for downloading; everything is automatically recorded from an online system's first message until logoff. Users don't have to strike any function or command keys to activate this capability. The program keeps the unprocessed downloaded file (in what SearchWorks calls the downfile) and the processed downloaded file (the outfile), so there is always a master copy to reformat later if errors are made on the first attempt (unless you fail to rename it, and overwrite the downfile with the next search).

It is best to issue the online print command for the records in as simple a document format as possible, excluding tags not pertinent to the user's needs (in BRS, for example, use . **.p5 au,ti,so,yr,lg,ab/1-5** and not . **.p5 all/1-5**). The only way to make prints during the course of the search is to do screen dumps as simultaneous printing and downloading onto a floppy is not possible during the online search.

FILE MANAGEMENT

File management functions include making a duplicate internal copy of a file, copying an internal file to an external medium, checking disk free space, truncating accounting records, and displaying, printing, erasing or renaming files. Editing consists of removing entire records and not specific fields or items within records from the downfile. In SearchWorks you choose records you want to delete, and in Pro-Search you choose records you want to save.

Processing files means ranking, sorting and/or indexing. Unprocessed files cannot be edited (this means that if you choose to not use the post-processing capabilities of SearchWorks you cannot remove unwanted records), and untagged files cannot be sorted. Another feature of the editor is the 'search-string' function, where records in a processed file can be searched online for the presence of keywords.

The maximum number of records which SearchWorks can process is 4,000. Up to thirteen different files from any combination of databases or online systems may be merged into one output file (each different file must, however, be processed separately and sequentially).

FIELD TAG TEMPLATES

The ability to translate tagged fields into common English and maintain a library of them is a very real and definite asset of SearchWorks. A template is limited to defining twenty-one field tags, must be defined for each file that is to be processed, and is used to provide a brief description of the contents of a downloaded file. Some databases have more than twenty-one tags (for example, **ERIC** has twenty-three and several business databases in DIALOG have over one hundred), and OnLine Research Systems is considering expanding the field tag template to one hundred fifty to accommodate these larger records.

To establish a template you first specify the file to be processed, name the template (usually the abbreviated name of the file; ERIC, for example), and the online system. This does not happen quickly at first because all three variables are not entered the same way and it is a little confusing. When this process is completed the top eight lines of the screen become a window into the downloaded file and you can scroll up and down to identify the appropriate record separator and tags for translation. The record separator, RS, is the first common tagged element of the search set. Identifying it alerts the software where individual records begin and the end of file indicator (the cost command) tells it where the file ends. Once templates have been established, accepted and saved, they may be recalled for future use. Processing new files from the same databases becomes much less cumbersome and time consuming. See Figure 3 for an example of an ERIC template.

RANKING, SORTING, AND INDEXING

Ranking or sorting (they're mutually exclusive; both cannot be done at the same time) are options which permit a reordering of the records from one or more files into a single file. The sorting option is chosen on the template and may be done on any one of the field tags which are common to all records.

Ranking involves an ordered weight sort of the words in records, from those with the most cumulative 'weight' to those with the least. Weight is user-specified on a template identical to that used for uploading. Up to ninety-nine term = weight algorithms may be entered with a numerical weight range of 1 (lowest) to 9 (highest). Words entered must begin with a letter (a-z), contain at least four characters, and exclude spaces plus any form of punctuation. The terms chosen should be unique and descriptive enough to allow one document to be distinguished from another. In a search on integrated pest management, for example, you would not want to rank management = 9 because it is not unique. Within the ranking you cannot be selective and specify where you'd prefer the words to be searched; it acts like a basic index search and looks everywhere for your term. The system allows for limited truncation, and it cannot search stop words (SearchWorks maintains an undocumented 50-60 item stopword list).

Both sorting and ranking reorder the original search set numbers and generate an index of ranked/sorted keywords which is printed with the processed search. The parameters for establishing an index include setting minimum and maximum occurrences of a word. These are identified in the profile and may be changed at any time. When presented with decreasing space on the RAM and a large file to index, SearchWorks will only index the longer terms. The index ignores stopwords, terms less than four characters in length, and terms (strings) which begin with numbers or consist only of numbers. Any punctuation between two words, like a hyphen, will result in the system splitting the word into two. Processing time for creating an index takes approximately 1.5 seconds per 1000 characters.

ACCOUNTING AND COVER SHEETS

The record keeping functions of SearchWorks are basic accounting functions which include summaries of amounts allotted, encumbered, unencumbered and net for nineteen different fund accounts, the ability to add, subtract, rename or display fund status, and a monthly/yearly report summary by online system accessed. Totals are automatically tallied and adjusted when any values are changed. The user profile contains general parameters which carry over to the cover sheet and serve as an invoice. Unlike Pro-Search, SearchWorks does not keep general statistics ►

on number of searches, searchers, databases accessed, or client expenses.

Information which appears on the cover sheet is date, patron, department, databases searched, search topic (three lines are allowed, but there isn't enough room to reproduce the search strategy here), results (this can be suppressed), amount and date due, search analyst, phone number and notes (three lines permitted; Pro-Search allows ten lines for notes). Not printing the cover sheet is also an option. SearchWorks provides a template for each search's cover sheet (two are produced) and automatically transfers information from your profile and from the accounting functions to the cover sheet; the remainder of the information is search-specific and must be entered for each search.

STRENGTHS AND WEAKNESSES OF SEARCHWORKS

SearchWorks is quickly and easily installed. Response time on a hard disk system is excellent, and the package is not difficult to learn to use. It is convenient because most responses in the menus are irrevocable; that is, when you are presented with a variety of choices you have only to type a letter for that particular operation to commence—you do not need to strike the return key. Highlights of SearchWorks' operation include effortless logon, automatic downloading, password masking, file merging (Pro-Search cannot), type ahead, search statement edit and recall, an online search guide, and certain aspects of its accounting function (in particular, the ability to add supplemental charges and service fees to the invoice). Deleting records, reformatting two-letter field tags, maintaining libraries of templates, uploads, downloads and reformatted files, ranking or sorting search results, generating indexes to searches and building small private databases all contribute to SearchWork's greatest strength: its ability to modify search results and improve the end-product for the user.

No software package is perfect and this package is no exception to that rule. SearchWorks can do a lot of work for you, but it does so according to its programmers' mindset, not yours. In short: you adapt to it, not it to you. Some minor modifications to your personal searching style and technique will be necessary (for example, entering correct system date and time to calibrate the cover sheet and requesting a cost command prior to database changes or system logoff to activate the accounting function). Portions of the manual and a few of the on-screen messages are not clear; there are small inconsistencies in portions of the menus, and the package itself does not have the visual amenities of Pro-Search. SearchWorks is awkward to run on a PC, and the accounting function it offers is basically weak. The limitation to 99 set lines in uploading will restrict excessively long search statements, and the lack of 'wraparound' in these line sets is frustrating. SearchWorks does not support printing or downloading to a patron's disk during the course of an online search. These activities are only possible *after* system logoff. The program's inability to provide printing simultaneously with the search is irritating, as is the incapability to selectively toggle either function on or off (as in SmartCom II). It is also important to realize that downloaded files which are too large for a client's disk cannot be divided and transferred piecemeal.

You cannot really fault a program fo: not doing something it was not designed to, but SearchWorks would be more useful if it could support multiple passwords for multiple online systems, provide assistance with database selection and use, translate the syntax of one search language to another, and offer editing within records. You cannot, for example, enter text such as local locations or call numbers, nor can you remove specific characters or fields within records.

RECOMMENDATION

SearchWorks has definite utility and is one of the better online searching aids presently available. Its capabilities overlap with, yet are dissimilar to, other available packages. It is not perfect, but it does complete the functions for which it was designed very well. It does not have the strength, versatility, or cost (!) of the three-part systems of ISI's Sci-Mate Searcher, Manager, and Editor ($999.00), PBS's Searcher's Tool Kit (Pro-Search, Pro-Cite and Biblio-Links) ($995.00), nor the combined powers of separate communications and word-processing packages. SearchWorks is a very credible and useful product which is uncomplicated and easy-to-use. It represents the first generation of online searching aids which incorporates some word processing capabilities into *one* convenient and affordable searching package.

This product fills a gap in the marketplace. It would be most useful for those searchers with a hard disk who have neither the time nor the money to purchase and maintain a working expertise with other more expensive and complicated systems. Searchers with the aptitude to balance the contradictions and stylistic differences of a communications package with a separate editing package, such as INSTANTCOM with PC-Write (see Miller's article in the March 1985 *ONLINE*) or SmartCom II with WordPerfect, would want Search-

Works only for its ability to rank keywords and assign weights to pertinent references. They probably would not buy it solely for its communication or searching capabilities. If you do not print most of your search results online, you may want to consider a package other than SearchWorks since that would negate the primary purpose and value of SearchWorks which is its ability to post-process online search results.

SearchWorks can, however, stand alone and be used for all of your searching needs unless you frequently access online systems other than BRS, DIALOG, NLM or SDC. Since SearchWorks cannot store the logon protocols for these other systems, you may prefer to use Pro-Search or a communications package which has that ability.

SearchWorks has no overwhelming or serious flaws. It contains one irritating oversight, the inability to selectively download or print continuously while online. If you purchase it, you will have to make a few modifications, some useful, some annoying, in your personal searching habits and styles, but the benefits of the package outweigh the compromises it requires.

Overall, SearchWorks is an excellent return for the money invested. It should not sit on your shelf gathering dust as poorly written and difficult-to-use packages do. SearchWorks is an apt name. It actually does work.

PRODUCT DESCRIPTION

Title:	SearchWorks
Audience:	Experienced searchers of intermediate or expert ability; not for the neophyte or end-user; most appropriate for those with a working familiarity with personal computers.
Producer:	OnLine Research Systems, Inc., 2901 Broadway, Suite 154, New York, NY 10025
Phone:	Ordering information: (212)-408-3311 Support Desk (1-5pm): (212)-249-4308
Contact:	Mark Nelson
Prices:	$199.00 each, discounts available for purchase of 5 or more; shipping costs $10.00; prepayment preferred
Updates:	A time schedule and price of software enhancements, revisions to the manual and alterations to the online search guide has not been announced.
Availability:	Direct from producer.
Online Systems:	BRS, DIALOG-2, SDC, NLM; can serve as a link to any other online system but only as a communications and downloading package. (Post-search processing isn't possible, and online search guides don't exist for any other online services).
Special/ Unique Features:	Post-search processing, ranking or sorting of search results, index generation, templates defining two-letter field tags.
Standard Features:	Automatic logon, uploading, type ahead, online search assistance, accounting, cover sheets.
Documentation:	Manual and quick-reference pages
Hardware:	Microcomputers: IBM XT, IBM AT, IBM PC or 100% compatibles; one double-sided and one fixed disk, or two double-sided floppy disk drives; it performs best on a hard disk. Operating System: DOS 2.0 or later Memory Requirements: 256K Modem: Hayes compatible, 1200 or 2400 baud Storage: Supports 5 1/4 floppy disks or fixed disks Printer and monitor
Software:	Two operating diskettes
Copyrighted:	Yes

►

FIGURE 1

SEARCHWORKS MAIN MENU SELECTION

```
a  –  Edit search strategy
b  –  Call
c  –  Process search results
d  –  File management
e  –  Print a file
f  –  Accounting & reports
g  –  Record editor
h  –  Modify profile
i  –  Help
j  –  Exit
```

FIGURE 2

FUNCTION KEY DEFINITIONS

```
F1    –  Online manual
F2    –  Scroll downloaded file & scroll lock (toggles)
F3    –  Clear downloaded file
F4    –  Mark manual line for display
F5    –  Display function keys
F6    –  Upload file and upload interrupt (toggles)
F7    –  Respond "y" to NLM/SDC "CONTINUE PRINTING" query
         (toggles)
F8    –  Type ahead of host (toggles)
F9    –  Send Break signal to host
F10   –  Disconnect from network

         Alternate function keys. Depress "Alt" and hit key.
F3    –  Open new file for downloading
F5    –  Display editing keys
F6    –  Transmit entire file to host (without pause)
F7    –  Check disk free space
F8    –  Set undefined prompt (toggles)
F9    –  Flush receive buffer ("unhang")
```

FIGURE 3

TEMPLATE FOR ERIC FIELD TAGS WITH A TOP WINDOW INTO THE SEARCH FOR BROWSING

```
LG  EN. .
IS  CIJNOV85.
CH  IR514394.
PT  080: 141.
AV  UMI.
YR  85.
MJ  Computer-Software. Databases. Information-Systems.
    Microcomputers. Online-Systems.
________________________________________________________

     File: gateway     Template: eric     Vendor: BRS
________________________________________________________

     RS Search/Replace: AN = ERIC Accession Number
     EOF Marker          : EST
      •^AU  =  Author(s) of Publication
       •TI  =  Title of Article or Document
       •SO  =  Journal Name
       •LG  =  Language of Publication
       •IS  =  Issue of Index
       •CH  =  Clearinghouse Code
       •IN  =  Institutional Affiliation
       •PT  =  Publication Type Code
       •YR  =  Year of Publication
       •MJ  =  Major Subject Descriptors
       •MN  =  Minor Subject Descriptors
       •ID  =  Identifiers
       •AB  =  Descriptive Abstract Summary
       •AV  =  Availability
       •NO  =  Grant/Project/Report Numbers
       •PR  =  ERIC Price Codes
       •LV  =  Availability Level
       •NT  =  Notes
       •GS  =  Geographic Source
       •
________________________________________________________
```

Put caret (>) at beginning of field to be sorted

FIGURE 4

RECORD FROM A PROCESSED SEARCH

Note the natural language field tags, which were derived from the template in Figure 3.

```
–01–
    ERIC Accession Number
        ED251620.
    Author(s) of Publication
        Aagard, James A.; Ansbro, Thomas M.
    Institutional Affiliation
        Naval Training Analysis and Evaluation Group. Orlando, Fla.
        BBB17048

    Title of Article or Document
        The Naval Enlisted Professional Development Information
        System (NEPDIS): Front End Analysis (FEA) Process. Technical
        Report 159.
    Language of Publication
        EN. .
    Geographic Source
        U.S. Florida.
    Issue of Index
        RIEMAY85.
    Grant/Project/Report Numbers
        RN: TAEG-TR-159.
    Clearinghouse Code
        CE040189.
        GV Federal.
    ERIC Price Codes
        EDRS Price – MF01/PC02 Plus Postage.
    Publication Type Code
        143.
    Availability Level
        1.
    Notes
        47o. : For a related document, see ED 218 476.
    Year of Publication
        84.
    Major Subject Descriptors
        Adult-Education. Employment-Qualifications. Job-Analysis.
        Military-Training. Occupational-Information.
    Minor Subject Descriptors
        Behavioral-Objectives. Career-Development. Databases.
        Educational-Needs. Employment-Opportunities. Job-Skills.
        Military-Personnel. Postsecondary-Education. Program-
        Design. Program-Development.
    Identifiers
        IDENTIFIERS: Front End Analysis. Navy.
    Descriptive Abstract Summary
        The Naval Enlisted Professional Development Information
        System (NEPDIS) was designed to function as a fully com-
        puterized information assembly and analysis system to sup-
        port labor force, personnel, and training management. The
        NEPDIS comprises separate training development, instruc-
        tional, training record and evaluation, career development, and
        audit subsystems. This report focuses on the . . .
```

Communications to the author should be addressed to Bill Coons, Albert R. Mann Library, Cornell University, Ithaca, NY 14853-4301; (607) 255-7731.

Communications Costs For Database Access—Waiting Is Costly

by Ralph Lee Scott
Associate Professor
East Carolina University

If you have ever had that "I've been sitting here a long time," or "this network seems slower than the other one" feeling, you are probably right! There is a difference in access time and consequently, the cost, between the four major United States packet switching networks (DIALNET, TYMNET, Telenet, and UNINET). Most database searchers, both professional intermediaries and end-users, call on these networks to transmit data between their terminals and a databank such as DIALOG, BRS, SDC Orbit, and others. While the pennies cited in this article do not amount to a great deal of money on the test searches shown, if you are online many hours a month, as most of us are, the cost differentials can add up. This article is divided into two parts. The first part, a brief history of the U.S. domestic packet switching networks, is followed by the practical results of this comparison project, including methodology and a statistical analysis.

HISTORY

The first nationwide domestic packet switching network was developed for the Department of Defense in 1969 and was called Arapanet. Some six years later (in 1975), the first of the domestic public packet switching networks was introduced by the Telenet Computer Corporation. The Telenet network enabled computer users across the nation to transfer information back and forth via the public switched telephone network (i.e., the phone system). Telenet was followed a year later in 1976 by TYMNET, Inc., of Cupertino, California, and by UNINET, a part of United Telecommunications, in 1980. A decade after its founding, TYMNET had 2100 access points, and 4300 communication links in 65 countries and 550 cities. Daily traffic on the network included some 530,000 customer transactions, sending some 87 billion characters over a one month period. TYMNET, just one of the packet switching networks, can alone handle over 6500 users at one time. The latest of the domestic packet switching networks reviewed here, DIALNET, a part of the Lockheed Information DIALOG system, began operations in 1984.[1]

PRIVATE AND PUBLIC NETWORKS

Private data communications networks, such as those used by OCLC and SOLINET, provide permanently wired long distance connections between terminals and computers using what is called a dedicated line. This means that only one type of data transmission can occur on that line at a time (unless the line is multiplexed). The public dial up packet networks, on the other hand, allow different data from many users to be transmitted between various cities. This is a much more cost-effective method ensuring that one long distance telephone line is utilized to its maximum extent.

In the public dial up networks, different users are gathered together at a local access point called a node. These nodes are reached in the local city by calling a regular telephone number. Examples of public dial up networks in the U.S. include: TYMNET, UNINET, DIALNET, and Telenet. After calling the local node, the data is transmitted by the public packet network to another node where the receiving terminal or computer is similarly connected to the network. There are standards used by these networks to enable the different terminals and computers to talk to each other. The most common current U.S. standard is RS232C. There is another standard, an international one, that is gaining domestic usage called X.25. These packet switching networks can operate at high speeds of up to 1.5mbps, but the average user on the local, in-town level, is often limited to 2400, 1200 or 300 bps.

FOUR NETWORKS COMPARED

This study compares the four main domestic dial up packet switching networks: Telenet, DIALNET, TYMNET and UNINET. Each of the networks uses a slightly different method of arranging and transmitting the packets. Factors affecting the speed in handling the packets are: network packet handling efficiency, the number of users at each of the nodes, the response time of the databank's computer, and rank of the packet in the status queue. The last item becomes important during national quasi-emergencies (such as the Grenada invasion, or a space shuttle liftoff), when military and governmental network needs strain the system capacity. At times like these, libraries sit around waiting. Presumably, the generals do not.

---

[1] Rajaraman, A. S., "Topological considerations in Tymnet." *Telecommunications* v.19., no. 9, (September 1985) pp. 56a-56g.

For the purpose of this study, a test search was formulated using two search terms in the **Ontap ERIC (File 201)** database on the DIALOG Information System. The terms were combined and printed out in bibliographic citation and full record format. These searches were then run as quickly as possible on the four public packet switching networks (UNINET, DIALNET, Telenet, and TYMNET). Nodes in New York and Los Angeles were accessed on four different dates (September 27th, October 15th, 19th and 20th, 1985). September 27th was the day Hurricane Gloria made her way up the East coast of the United States. Interestingly, the storm had little impact on transmission times on the transcontinental domestic telephone network. October 19th was a Saturday and the 20th was a Sunday. End-users who do searching at home or at work on the weekends or during evening hours, will find that the cost differentials on nights and weekends, when the networks are more stable, will be less. It is to your advantage to do your searching during these stable periods. You will note on the graphs that not only is the cost and time on a given network more stable (say DIALNET for instance), but the differences *between each of the networks* is also less on nights and weekends. The searches were conducted at random during morning, afternoon, and evening Eastern Daylight time periods. Forty eight searches were conducted, twelve on each of the four networks. The local nodes in New York and Los Angeles were accessed by dial up telephone service (via Carolina Telephone and Telegraph Co. and the AT&T System) from Greenville, N.C., the author's residence. The transmission time and costs to each of the four network nodes were considered uniform for a given city. The cost and time measured was that from the network node to the DIALOG computer. Every effort was made to ensure that the forty eight searches were entered and run at a uniform and rapid operator rate. If the searcher had a problem, or an error occurred, the search was run over. The Ontap ERIC database was selected because it was cheap, and provided a small file with quick look up/response time.

TEST RESULTS

The mean time to look up two terms, combine them and print out the seven bibliographic citations (DIALOG format 3) was 0.0097916 hours. The mean cost was 23.39 cents for all times, cities, and networks. To print out the same seven citations in full record format (format 5) required a mean of .03029166 hours and a mean cost of 72.83 cents.

The hourly connect charges for the searches done in October on the four networks were as follows: TYMNET, UNINET and Telenet, each $10/hr. and DIALNET, $6/hr. The time required to run the sample search in format three ranged from a low of .007 hours under DIALNET to .025 hours for one Telenet search. For format three the mean cost was 23.395 cents. The mean cost for all cities and dates in format three was as follows: TYMNET, 22.42 cents; Telenet 29.66 cents; UNINET 25.08 cents; and DIALNET 16.75 cents. The deviation by city in cost was slight.

The cost for printing out all of the records retrieved in format five (full record on DIALOG) varied in a similar manner as the format three types. With a mean cost of 72.83 cents for all searches, the network means varied as follows in format five: TYMNET, 67.50 cents; Telenet, 92.08; UNINET, 77.50 cents; and DIALNET, 54.00 cents. As was the case with format three types, the city variation was slight.

TECHNIQUES DIFFER

The three major packet switching networks use different methods, controls and instructions for packet processing. These procedures are known as flow control strategies. They have as their objective the prevention of throughput degradation, deadlock avoidance (where the network chokes to death on its own traffic), fair resource sharing, and protocol matching between the user and the network. The efficiency with which these strategies operate control the end-user's cost. As we have seen in this study, there is a variation in the mean network cost between UNINET, DIALNET, TYMNET and Telenet. It is postulated that these costs are related to the speed with which the three networks control their data packets.[2]

CONCLUSION

If it costs more to use Telenet and less to use DIALNET, why it does, while useful to know, is not essential to making the best use of funds available. Perhaps DIALNET is cheaper because there are a smaller number of users on the system. The point of this study is: *don't stick to the same network out of habit.* You may have to learn a slightly different network protocol, but with several hundred annual hours of searching, the cost between a 54 cent search and a 92 cent one can mount up. For example, 1000 searches run at the mean DIALNET cost would come to $540.00 while the same number of searches run on Telenet would come to $920.80. The use of DIALNET would lead to a saving of $380.80 or 41.33 percent. It is hoped that this study can be repeated in future years. In the meantime. . .watch those costs on your favorite network! You *are* sitting there a long time!

---

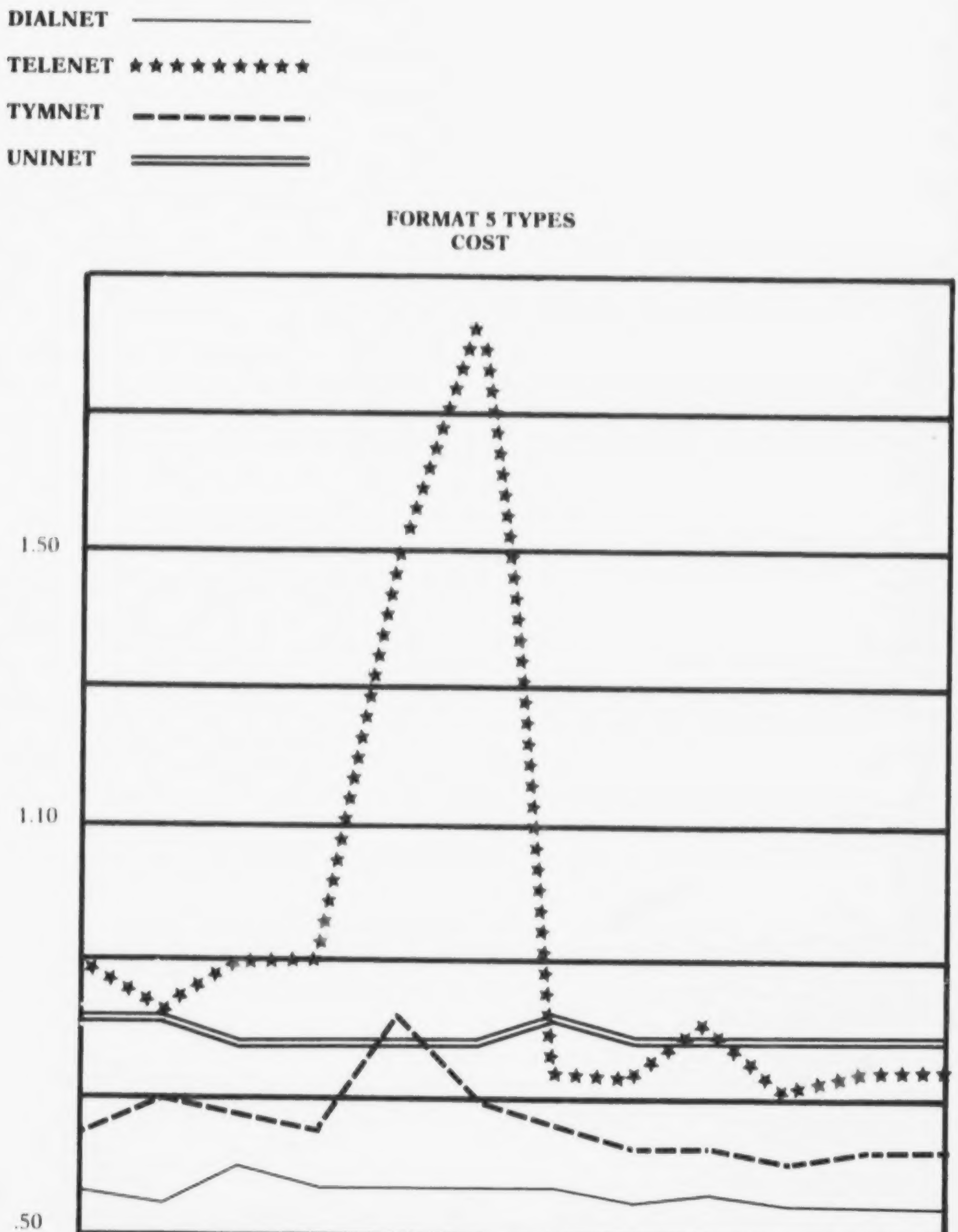
[2] Hsieh, W., "Performance analysis of an end to end flow control mechanism in a packet switching network." *Journal of Telecommunications Networks* v.2., no. 1(Spring 1983) p. 103.

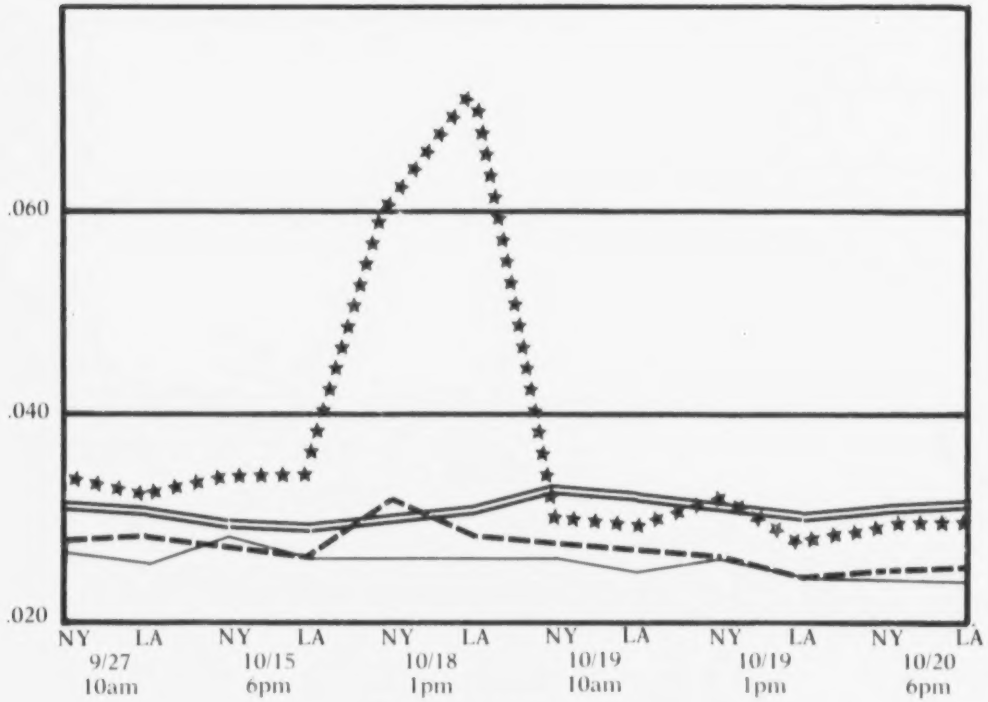
TABLE 1
Mean Cost

	DIALNET	TELENET	TYMNET	UNINET
Format 3	.17	.30	.22	.25
Format 5	.54	.92	.67	.77

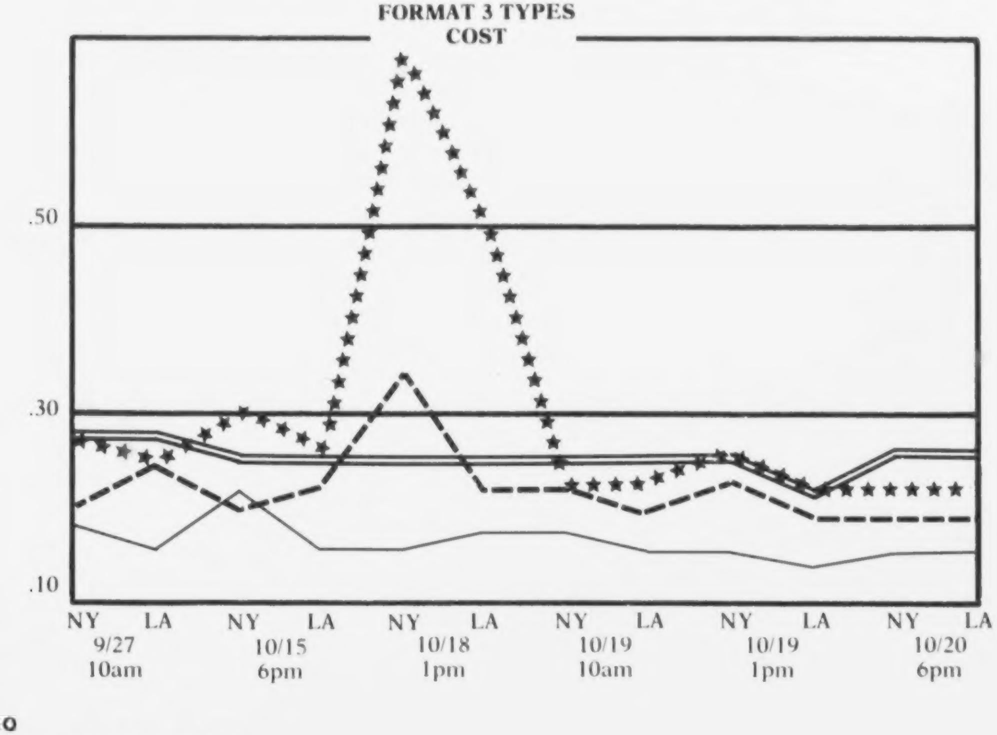
FIGURE 1
Sample Printout With Costs

```
 5/3/1
EJ111324        IR501421
    Mollie Huston Lee: A Profile
    Moore, Ray Nicholas
    Wilson Library Bulletin, 49, 6, 432-9 Feb 1975

 5/3/2
EJ107658        IR501243
    Observing World Population Year; Films for Thought
    Cleaver, Betty
    Previews, 3, 4, 3-5 Dec 1974

 5/3/3
EJ106216        IR501174
    The De-Traditionalizing of English 101
    Richardson, Donald
    Learning Today, 7, 4, 55-62 F 1974

 5/3/4
EJ106197        IR501155
    Radio-Free Library Power
    McRae, Jane Carroll
    Wilson Library Bulletin, 49, 3, 224-8 Nov 1974

 5/3/5
ED107260        IR002000
    Learning Materials and Services at the University of North Carolina
      at Charlotte.
    North Carolina Univ., Charlotte, J. Murrey Atkins Library.
    Oct 1974 43p.
    EDRS Price – MF01/PC02 Plus Postage.

 5/3/6
ED098964        IR001343
    Directions and Shortcuts to Book Ordering. Second Revised Edition.
    Schaberg, M. P., Ed.
    North Carolina Univ., Chapel Hill. Library.
    1973 17p.
    EDRS Price – MF01/PC01 Plus Postage.

 5/3/7
ED096970        IR001150
    Evaluation Report for Institute for Public Libraries in Service to
      Young Children, 1972-1973.
    Learning Inst. of North Carolina, Durham.
    Sep 1973 72p.
    Sponsoring Agency: Bureau of Libraries and Educational Tech-
      nology (DHEW/OE), Washington, D.C.
    EDRS Price – MF01/PC03 Plus Postage.
?cost
                 19oct85 12:27:15 User004799
$0.10      0.007 Hrs File201
          $0.00 7 Types in Format 3
$0.00     7 Types
$0.04     DIALNET
$0.14     Estimated cost this file
$0.15     Estimated total session cost   0.008 Hrs.
```


DIALNET
TELENET
TYMNET
UNINET
FORMAT 5 TYPES
COST
1.50
1.10
.50
NY
LA
9/27
10am
NY
LA
10/15
6pm
NY
LA
10/18
1pm
NY
LA
10/19
10am
NY
LA
10/19
1pm
NY
LA
10/20
6pm

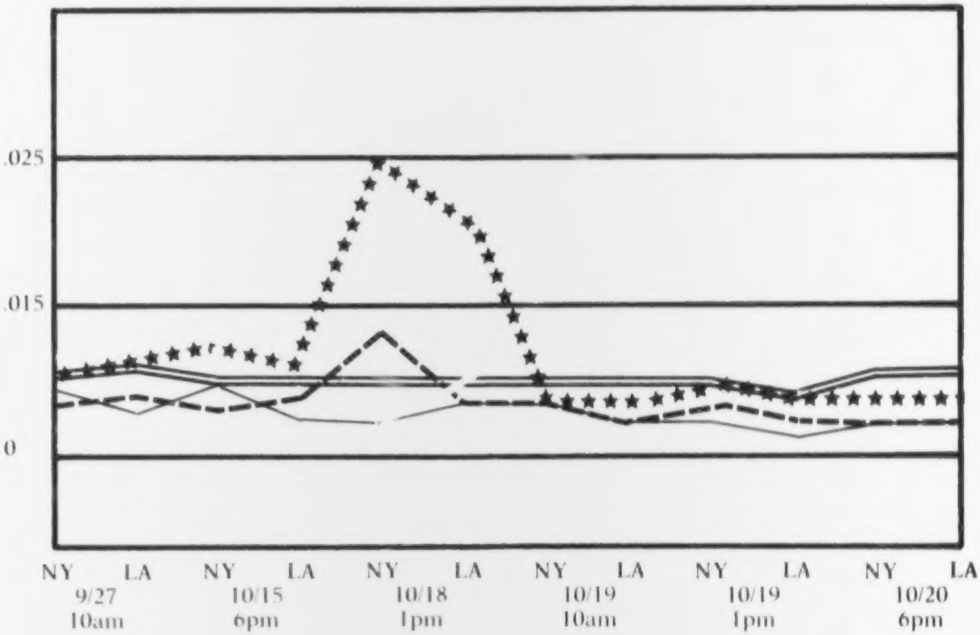


FORMAT 5 TYPES
TIME



FORMAT 3 TYPES
COST

FORMAT 3 TYPES
TIME



THE AUTHOR

RALPH LEE SCOTT has a BA in History and a MS in Library Science from Columbia University and a subject MA in History from East Carolina University. He is currently Associate Professor of Academic Library Sciences and Head, Government Documents/North Carolina Unit at East Carolina University's J. Y. Joyner Library. His interests in computer-based technology date back to 1968, when he designed a prototype indexing scheme that ran on the IBM 360/91-75 for incunabula based on the De Ricci Census. He has been searching databases, first offline through the North Carolina Science and Technology Research Center since 1972 and then online through the DIALOG system since 1976 at East Carolina. His interest in the Domestic Packet Switching Networks, dates from his many hours spent waiting for the prompt to come back.

Communications to the author should be addressed to Ralph Lee Scott, Associate Professor, Academic Library Services, J.Y. Joyner Library, East Carolina University, Greenville, NC 27834.

WHAT'S IN A NAME?

USE OF NAMES AND TITLES IN SUBJECT SEARCHING

by Anne B. Piternick
Professor
University of British Columbia

INTRODUCTION

Subject searches of bibliographic and source databases using names of persons or organizations, or titles of books or movies, are usually carried out because the searcher wants information about these individuals or works. Searching may be favored in some databases by the provision of special fields, such as "Named Person" or "Document Type."

Names and titles may, however, also constitute useful search terms for information on topics rather than on the person or work named. Persons or organizations, or even fictional characters, often come to represent particular concepts or schools of thought with which they have been closely associated; published works become metaphors for the ideas or principles they expound. So we develop eponymous terms and adjectival phrases such as "Marxism"; "Freudian psychology"; "Bauhaus" style; "in a Pickwickian sense"; "a Catch-22 situation"; "the Third Wave." The "China syndrome," originally a term used by nuclear scientists, was popularized by its use as the title of a movie.

> **Persons or organizations, or even fictional characters, often come to represent particular concepts or schools of thought with which they have been closely associated; published works become metaphors for the ideas or principles they expound.**

The use of names and titles is particularly helpful when we wish to express abstract concepts, or concepts which are novel. This article looks at some ways in which names and titles may be used in online searching, especially in defining new concepts, in enhancing searches by improving recall or precision, and in developing search expressions for multi-database searching. The following categories will be considered: personal names; titles of books and movies; names of affiliating and sponsoring bodies; and journal titles.

PERSONAL NAMES AS SUBJECTS

Citation searching has clearly demonstrated the effectiveness of using names of individual authors and their works in subject retrieval. The theory behind citation searching is that the author's publications can be used to represent the topics on which he or she has written. Thus, a later paper citing a particular publication is likely to deal to a greater or lesser extent with some aspect of the topic treated in the earlier work. Especially in co-citation searches of specialized subject areas, names alone, without citations to individual published works, may represent bodies of writings by the authors, or what White has called "oeuvres," and thus be sufficient to define the subject areas in which the authors worked.[1]

This same principle lies behind the use of personal names in the development of search strategies using natural language free-text terms. We have in recent years seen numerous examples of medical or medico-legal cases where the person or persons involved have come to stand for unorthodox experimentation, or situations involving medical ethics. "Rent-a-womb," "right-to-die" or "right-to-life" cases bring with them dilemmas which have been widely discussed outside the medical and legal communities. Free-text searching of the name (or pseudonym) of the individual in question is a rapid way of identifying terms used to index items on such controversies, especially in searching newspapers and popular magazines. Recent "transplant" examples are Barney Clark, William Schroeder and Baby Fae. Knapp notes inclusion of an entry in the **TERM** database for "Baby Jane Doe," identifying free-text, **MeSH**, and **PsycINFO** terms which focus on the right to life of newborn infants with severe abnormalities.[2]

One area where the use of personal names is of particular value is Law. Legal cases, especially landmark cases, become known by the personal or case name of the litigant, and thus this name becomes synonymous with the legal principle or concept involved: Miranda with rights of a person charged with a crime; Hinckley with a re-examination of the plea of insanity. Since landmark cases are normally of importance because they establish new legal concepts or principles, it may be some time before a standard term is developed to describe such concepts or principles. In the interim, the case name provides an alternative subject descriptor, and may continue to be useful as a free-text term after an "official" term has been coined. The case name may, in fact, have a long life as a lay term, and thus be especially useful for searches of non-legal databases.

One such case was that of Bakke, the first publicized case in the United States to deal with the concept of "reverse discrimination." A search carried out in late 1982 for articles on this concept showed the usefulness of the name as search term. Articles in the **National Newspaper Index (NNI)** and **Magazine Index** discussing impacts of the Bakke decision were searched on DIALOG using BAKKE as a named person **(BAKKE/NA)** or free-text term. Most relevant items retrieved had been assigned the descriptor *DISCRIMINATION IN EDUCATION-CASES*—quite a broad term. It became apparent, however, that indexers were beginning to recognize the need for a more specific term to use for this new concept when records were retrieved in which "Bakke" was used as a title enhancement, and *Bakke case* as an identifier, for documents which were not dealing specifically with the Bakke case as such. Using "Bakke," references were also retrieved to the related Weber case, and, later, records showing that the new term *reverse discrimination* had finally come into use.

Negative use of "Bakke" in the search **REVERSE ( )DISCRIMINATION NOT BAKKE** turned up a number of articles on the principle, especially as applied in employment in general, to women, and so on, without specific reference to the Bakke case.

This search also illustrated some of the dangers of personal name searching. A search of the **PAIS** database turned up a book on shipping which had a foreword by one Karl E. Bakke, and one item retrieved from the Magazine Index using **BAKKE/NA** turned out to be an interview of Horst Bulau by Bill Bakke about ski-jumping. There were few such false drops with a name as uncommon as "Bakke," but with common names in particular other terms might be added to the search expression to avoid problems of ambiguity—for example:

BAKKE AND DISCRIM? .

►

Such a procedure has been shown to work well when searching acronyms.[3]

A name used in subject searching need not be that of a real person. Dadlez used the name of a literary character, "Big Brother" as a metaphor carrying all the semantic connotations of Orwell's novel, *1984*. She notes that:

> "When the phrase "Big Brother" is used outside the context of the book, it invites a plethora of analogies, perhaps even more so because it allows for the personification of abstractions and provides a single image as a vehicle to carry associations."[4]

Dadlez also discovered, after an initial search on a large number of BRS databases, that limits need to be placed on the term so as to avoid irrelevant associations: a refined search excluded terms relating to family or siblings or the Big Brother Program. Dadlez's article is interesting in that she conducted a thematic analysis of her search results, and identified in them four main themes relating to the "Big Brother" metaphor.[5]

DOCUMENTS AS SUBJECTS – TITLES OF BOOKS AND MOVIES

Dadlez's article is one of many prompted by the significance of Orwell's book for the year 1984, when the concepts suggested by the title of his book were endlessly discussed with relevance to contemporary society. Other books and documents carry connotations which may be useful in searching. President Reagan not surprisingly chose a film when he tried to find an adjective to describe his high-tech, space-oriented defense policy: he called it "Star Wars."

STAR()WARS proved to be a useful term to use to plan a search for early items relating to the President's proposals and the implications of such a military policy[*]. Indexers obviously had some difficulty in selecting one or two controlled terms which would express this new concept. A search of PAIS using **STAR( )WARS AND DEFEN?** on DIALOG produced records which included an assortment of assigned descriptors:

ATOMIC WARFARE – STRATEGY
GUIDED MISSILES
GUIDED MISSILES – DEFENSES
SATELLITES, ARTIFICIAL – MILITARY APPLICATIONS
SPACE, OUTER – MILITARY ASPECTS
UNITED STATES – DEFENSES

A comparison search based on time-consuming selection of PAIS controlled terms turned up only one additional relevant citation, which had been indexed under *ANTISATELLITE WEAPONS*. These searches clearly indicated that **STAR( )WARS** would have rapidly provided a useful set of candidate terms for a further search based on descriptors.

Comparison searches were also tried on the Magazine Index database. Unlike PAIS, this database does not limit the number of descriptors used to index a single document; this search turned up documents with as few as one and as many as 21 descriptors. A study of terms used to index sample relevant citations is therefore a useful preliminary to searching this database.

The comparison searches turned up relevant documents indexed under one or more of the following descriptors; almost all of these were retrieved using **STAR( )WARS AND DEFEN?**. (All would have been retrieved using **STAR( )WARS** alone – but so would a large number of irrelevant documents.)

AIR DEFENSES, MILITARY – PLANNING
ANTIMISSILE MISSILES – ANALYSIS
ANTIMISSILE MISSILES – DESIGN AND CONSTRUCTION
ANTIMISSILE MISSILES – TECHNOLOGICAL INNOVATIONS
ANTI-SATELLITE WEAPONS – PLANNING
ARMS CONTROL – TECHNOLOGICAL INNOVATIONS
ARTIFICIAL SATELLITES – MILITARY USE
ATOMIC WEAPONS – TECHNOLOGICAL INNOVATIONS
BALLISTIC MISSILES – MILITARY USE
GUIDED MISSILES – RESEARCH
GUIDED MISSILES – TECHNOLOGICAL INNOVATIONS
LASER BEAMS – MILITARY USE
LASERS – MILITARY USE
MILITARY ART AND SCIENCE – TECHNOLOGICAL INNOVATIONS

---

[*] *Some readers may wonder why* **"STRATEGIC()DEFENSE()INITIATIVE"** *was not preferred as a search term to* **"STAR()WARS"**. *The answer is that this name was generally used for the program when it was first announced. The earliest citations produced by searches of* NNI *and* Magazine Index *using this search term or* **"SDI AND DEFEN?"** *were dated March and May 1984 respectively, and were retrieved because* STRATEGIC DEFENSE INITIATIVE *was assigned to these records as a descriptor. A search on PAIS using the same strategy produced documents published 1984 or later. These were retrieved because the term appeared in title, title enhancement, or abstract; as of May 1985, it was not being used as a descriptor by PAIS.*

MILITARY POLICY – ANALYSIS
PRESIDENTS – MILITARY POLICY
SPACE WEAPONS – MILITARY USE
SPACE WEAPONS – PLANNING
UNITED STATES – NATIONAL SECURITY
WEAPONS SYSTEMS – PLANNING
WEAPONS SYSTEMS – TECHNOLOGICAL INNOVATIONS
and one identifier: *SPACE WEAPONS – INTERNATIONAL ASPECTS.*

The Magazine Index and NNI are both indexed using the same vocabulary (both are IAC databases) and searches of these databases provided some interesting findings on the indexing of the documents.

In many cases it would be prudent to search terms alone without subheadings, since application of the latter appeared to be rather erratic. For example, although the indexer may be guided by rules to differentiate between the subheadings *MILITARY USE* and *MILITARY APPLICATIONS*, the searcher may not be sure which to use in a particular case. Choices between sub-headings such as *PLANNING* or *ANALYSIS* as applied to policy, and, *RESEARCH* or *TECHNOLOGICAL INNOVATIONS* applied to weapon systems, could be difficult for the searcher, and such sub-headings may be most useful in limiting the size of a final set. Both *MILITARY ART AND SCIENCE – TECHNOLOGICAL INNOVATIONS* and *TECHNOLOGICAL INNOVATIONS – MILITARY ASPECTS* were found as descriptors for "Star Wars" concepts. It would also be difficult for a searcher to decide whether *LASERS* or *LASER BEAMS* would retrieve the most relevant citations.

The retrieved sets did distinguish nicely between documents relating to military policy before and after the President's "Star Wars" announcements by using *UNITED STATES – MILITARY POLICY* (or *MILITARY POLICY – UNITED STATES*) before, and *PRESIDENTS – MILITARY POLICY* after. The **STAR()WARS** strategy also indicated that searching *REAGAN, RONALD – MILITARY POLICY* or *REAGAN, RONALD – SPACE POLICY* in the "Named Person" field would retrieve a high number of relevant citations – which it did from both sets, although some items either had no named people, or named others rather than the President. (This provides another example of the usefulness of searching on personal names in cases where a particular concept has been closely identified with an individual.)

Late in 1984, IAC finally introduced *STAR WARS (WEAPONS SYSTEM)* as a descriptor, and items published from approximately July 1984 are indexed using this term.

In the free-text searches described here, the term **DEFEN?** was included in order to exclude items relating to the movie itself. This did eliminate some "Star Wars" items which were retrieved in the search which used only controlled descriptors. Neither PAIS nor Magazine Index permitted exclusion of document type as a field, which would have been considered as a way of eliminating movie reviews; the NNI search, however, permitted use of an "Article Type" field, and thus **NOT AT=REVIEW** was used, as well as the **DEFEN?** restriction. The Magazine Index search produced a couple of interesting false drops, neither of which referred to the movie. An article on investment in high-technology defense industries had the title: "Playing *star wars*: Fred Kittler picks his favorite high-tech *defense* stocks." And an article evaluating video games was indexed under a long list of product names including both "*Star Wars*: The Empire Strikes Back" and "*Defender*"!

AFFILIATING AND SPONSORING BODIES AS SUBJECTS

The names of organizations which are not "authors" of documents, but which are in some way connected with or responsible for their existence, are also potential subject indicators.

> **The names of organizations which are not "authors" of documents, but which are in some way connected with or responsible for their existence, are also potential subject indicators.**

Organizations to which authors are affiliated, or which have sponsored research being reported, may be used in this way, and can be of particular utility in searching databases which do not provide a controlled vocabulary.

One example of such a database is **IEC** – the **Directory of Federally-Supported Research at Canadian Universities**, available on the Canadian CAN/OLE search system. Subject searching of this database can only be done using keywords in the titles of projects; other search fields are author, affiliation, and sponsor (a Canadian government department or agency). The database covers projects described in both of Canada's official languages. A search of IEC attempted to find projects on the study of Canadian languages, including Canadian English ►

and Canadian French as well as of native languages, and regional, local and ethnic dialects. A variety of English and French keywords were selected, to express the concepts of language, linguistics, speech, dialect, and so on.

As far as possible, it was hoped to avoid projects on the teaching of language in Canada, and also projects relating to machine languages. The simplest way to do this was found to be by means of the use of affiliation as a subject code representing the "subject" of a university department or institute. Thus, by excluding **Comput*** or **Math*** or **Educ*** and their French-language equivalents in the affiliation field it was possible to avoid the difficulty of preparing a bilingual list of keywords to represent these subjects.

Names of sponsoring agencies were also used in a limited way to reduce noise in this search. The Natural Sciences and Engineering Research Council, for example, was excluded because of the strong possibility that it would only fund projects on machine languages, machine translation, and the like.

Both affiliation and sponsorship might have been used positively as well as negatively for certain types of searches on this database. For example, a search for work on native languages could have been carried out with a high degree of accuracy by searching on keywords representing language, and combining with terms in the affiliation field to cover Departments of Anthropology, Native Studies, Etudes Autochthones, and so on. Funding for such studies in Canada would almost certainly come from the Department of Indian and Northern Affairs or the Social Sciences and Humanities Research Council of Canada, and these could also be specified in the sponsor field on IEC. The use of these affiliation and sponsorship terms would eliminate the need for a keyword search involving the names of all possible native languages, not to mention all possible French and English spellings of these languages, which are not standardized.

Although the IEC database is likely to be of limited interest outside Canada, the techniques described above could be successfully applied in searches of other databases which have no controlled vocabulary—for example, the Corporate Source field in **SCISEARCH** and **SOCIAL SCISEARCH**. Corporate Source is included in the Basic Index for both these databases on DIALOG. However, searching Corporate Source separately to create a "subject" subset has advantages. To use the first example discussed here: searching a "language" set and excluding **COMPUT?** in the basic index would eliminate any computer studies of language along with computer languages. Searching the "language" set in the Basic Index and excluding **CS=COMPUT?** would eliminate studies coming from Departments of Computing Science or equivalent, but would retrieve studies of computer applications coming from other departments.

Although the ERIC Clearinghouses do not sponsor work in educational research, they do have a responsibility for selecting documents to go into the **ERIC** database, and thus in some sense they "sponsor" the documents which describe the work. Because of this selection process, the codes for Clearinghouses representing different areas within the overall subject of Education may be used as equivalents to classification codes. An example using the code for the Clearinghouse on Science, Mathematics, and Environmental Education to limit a search to this subset of the database is given in the DIALOG Database Chapter for ERIC.[6]

JOURNAL TITLES AS SUBJECTS

Jennings has noted that, for small, specialized databases at least, the database itself represents a subject parameter which may be "assumed" in a search.[7]

In the same way, a journal dealing with a specific subject area is in itself a "database" of articles on that subject. Since it is possible to search journal titles (usually in abbreviated form, and occasionally as CODEN or ISSNs) on most bibliographic databases, they should be considered as search terms especially to enhance recall, or to retrieve highly relevant citations.

An excellent example of the use of journal titles is provided by Gertrude Foreman in a study of the adequacy of the **PRE-MED** database on BRS for providing access to the literature on Nursing.[8] Since the purpose of PRE-MED is to provide early access to key journals in clinical medicine, nursing, and hospital administration, indexing is done as rapidly as possible in order to eliminate the time lag between publication and inclusion in **Index Medicus** and **MEDLINE** which is necessitated by the thorough indexing procedure applied by NLM. PRE-MED subject indexing is based on the title of the article as it appears in the table of contents of the journal; searchers may use the MeSH terms assigned by the indexers and/or keywords in titles.

The depth of indexing in such circumstances depends primarily on the completeness and specificity of the article title, and Foreman found that neither the assigned descriptors (including the MeSH sub-coding for "Nursing") nor the title words provided adequate retrieval of items of interest to nursing, or relevant to the point of view of the nursing profession. She did discover, however, that all but one of the nursing journals indexed in PRE-MED included the truncated term **NURS** as part of their abbreviated titles. She thus included **NURS** in the "source" paragraph as one parameter of her search to enhance retrieval.

Although keyword searches of journal title fields are not always possible, one may consider applying the principle of Foreman's technique to searches of other databases where subject access is restricted. For example, searches of SCISEARCH and SOCIAL SCISEARCH might be enhanced by the inclusion of journal titles. The documentation for these sources provides subject-classified lists of important source journals in their respective fields, which might be considered as "hedges" or SAVEd searches to limit retrieval to highly-relevant, high-profile journals. Or lists could be compiled from the subject categories in *Ulrich's International Periodicals Directory* (in print or online on DIALOG or BRS). *Ulrich* usefully supplies ISSNs, which can be used on certain databases.

BRS, by arrangement with ISI, has created Category Codes for each of the subject categories used by ISI to classify its journal listings. These codes are then assigned to records in which the individual journal titles appear as sources. Thus retrieval can be limited to specific groups of journal titles merely by including the category code(s) as a term or terms in the search. Use of these codes is discussed in detail by Janke.[9] The ISI *User's Guide* discusses the use of titles of small groups of "prime" journals to partition the SOCIAL SCISEARCH database to produce a basic set of items highly relevant to a broad subject area; this set can then be further restricted to retrieve specific document types such as reviews or bibliographies, or combined with title words to increase subject specificity.[10] Feduska suggests limiting current awareness searches of PRE-MED on BRS to journals held in the library, or to journals not held locally and thus not available for current scanning. She compiled saved searches very efficiently in this case using ISSNs rather than journal title abbreviations.[11]

The start of publication of a journal often marks the recognition by workers in the field that a new subject specialty has been established. The use of the title of such a journal as an "OR" term with other subject terms may ensure recall of items relevant to the subject specialty, especially if the latter is not yet well represented by the controlled vocabulary.

As an example, indexing of the *American Journal of Art Therapy* was examined on several databases. This journal began publication in 1961, indicating that "Art Therapy" was an established area of scholarly study by that time. However, the term *ART THERAPY* was not established as a controlled term on PsycINFO until 1973, and not until 1977 in the *ERIC Thesaurus*. ERIC established *ART* as a controlled term in 1966 and *THERAPY* in 1969, and included *ART THERAPY* as an identifier (or candidate descriptor) in some 1976 citations. ERIC did, however, re-index many "ART AND THERAPY" citations retrospectively under *ART THERAPY*; had this not been done, the journal title search would have been useful for retrieving some early citations on ERIC.

A further examination of the indexing of the journal was made on the PsycINFO database on DIALOG. A total of 22 hits were retrieved using the formula **JN = AMERICAN JOURNAL OF ART THERAPY**, and excluding any cases where ART THERAPY occurred as a free-text keyword (**ART( )THERAP?**), or as a descriptor (**ART THERAPY/DE**). A number of these could have been relevant to searchers interested in Art Therapy. Several were indexed before *ART THERAPY* was established as a descriptor by PsycINFO, although the words "art therapy" did occur in identifier phrases before this time. Titles of interest indexed from 1973 on, included items on "nonverbal psychotherapy methods," "relation of art to psychotherapy," "activities therapy program" (for hospitalized patients), and "therapeutic technique using music and exercises and painting and drama."* These examples indicate that the journal title itself could have been used to increase recall of articles relating to the overall concept.

Journal titles may of course, be used to limit a set by excluding them from a search. As an extreme example, one researcher was looking for new applications for a chemical substance. He prepared a list of around 500 journals which would be likely to report known applications—in agriculture, medicine, pharmacology, metallurgy, and so on. He then formulated an SDI search expression which comprised the various names by which the substance was designated, **AND NOT** the long list of journal CODEN; the search was then regularly run on several databases. I am sorry I am unable to report whether or not he ever found a novel application for his chemical substance in this way.

One difficulty of using journal titles as subject indicators, especially in multiple database searches, is the lack of standardization in the way in which they are represented in different databases, and failure to use ISSNs and/or CODEN.[12]

CONCLUSIONS

Names and titles may serve as useful search terms for online searching. Their use should be

---

*Quotes are from identifier phrases or titles.

►

considered especially in representing new topics, for which generally accepted terms have not been—and may never be—established; where a name represents a popular or layman's approach to a topic; or where a topic is difficult to define because it is diffuse, especially in terms of a controlled vocabulary. Because they are highly specific, names or titles are especially useful in planning searches, when the searcher wishes to retrieve a few citations which will illustrate how a topic has been indexed in a particular database.

Names of sponsors or affiliations may provide an approach similar to that supplied by classification codes. Groups of journal titles offer a means of restricting searches to highly relevant journals important in their fields, and have potential as "hedges" or SAVEd searches for use on a recurring basis. Selected titles may be employed to increase retrieval of relevant items, or to exclude the irrelevant, especially in searching databases which have no controlled vocabulary.

Problems in using names or titles arise from inconsistencies in their representation, especially in full text, and in newspaper databases. Personal names are often reduced to surnames, and should be linked to other terms which are most likely to occur in the desired context. The lack of standardization of journal titles increases the work involved in compiling search expressions based on them, especially if several databases are to be searched. Similarly, representation and searching of names of affiliations and sponsors vary from database to database and from system to system. Finally, inclusion of a journal title in a list of serials indexed is no guarantee that items from that journal of potential interest to the searcher will have been picked up by the indexer.

In spite of these caveats, the name/title approach can provide a promising strategy in the kinds of situations indicated above. In order to enhance the benefits of this approach, the following recommendations are made:

- Where names or titles usefully represent concepts for which no preferred terms exist, indexers should consider including them as identifiers if this field is present in the record, or as title enhancements.
- Compilers of thesauri and other controlled vocabularies should consider including relevant names or titles as entry terms leading to preferred terms.
- Database producers and searching systems should cooperate in providing search fields based on ISSNs and/or CODEN in bibliographic databases which cover the journal literature.

ACKNOWLEDGEMENTS

Several of my students contributed, directly or indirectly, to this paper. I am grateful to Cynthia John, who conducted the Bakke searches; Deborah Scott-Douglas and Paola Durando who experimented with "Star Wars"; and Parker O'Brian, whose work suggested the journal search for articles on "Art Therapy."

REFERENCES

1. H. D. White, "A Cocitation Map of the Social Indicators Movement." *Journal of the American Society for Information Science* 34, 1983, pp. 307-312.
2. S. D. Knapp, "Creating BRS/TERM, a Vocabulary Database for Searchers." *Database* 7, December 1984, pp. 70-75.
3. C. W. Holt, "Online Searching with Acronyms" *Online* 7, November 1983, pp. 70-76.
4. E. M. Dadlez, "Orwell's *1984*: natural language searching and the contemporary metaphor." *Database* 8, June 1984, pp. 98-104.
5. Ibid, Table II.
6. Database Chapter for DIALOG File 1—ERIC. Palo Alto, CA: DIALOG Information Services Inc., Revised April 1981; pp. 1-13.
7. P. G. Jennings, [Letter to the Editor] *Online Review* 6, 1982, p.267.
8. G. Foreman, "Subject Access to PRE-MED Nursing Journals." *BRS BRIEF PAPER #28*. [Distributed with *BRS Bulletin* for May 1983].
9. R. V. Janke, "Searching the Social Sciences Citation Index on BRS." *Database* 3, June 1980, pp. 19-45.
10. *User's Guide to Online Searching of SciSearch and Social SciSearch*. Philadelphia: ISI, Revised 1980; pp. 8-12.
11. M. A. Feduska, "The ISSN and the Permanent Save: the Creme de la PREM." PREM & SPIF Searching Contest: Winning Paper for PRE-MED First Place. [Distributed with *BRS Bulletin* for January 1983.]
12. M. E. Williams and L. Lannom, "Lack of Standardization of the Journal Title Data Element in Databases." *Journal of the American Society for Information Science* 32, 1981, pp. 229-233.

Communications to the author should be addressed to Anne B. Piternick, Professor, School of Library, Archival and Information Studies, University of British Columbia, Vancouver, B.C. Canada V6T 1Y3

"Hedges"... The Sometimes Ignored Search Technique That Can Save A Lot of Time

by Genevieve N. Schiffmann
Reference and Bibliographic Services Section
National Institutes of Health Library

The National Institutes of Health Library has a very active literature search service which responds to requests from the whole spectrum of the NIH community of research scientists, patient-care staff, program/project-officer/analyst, and administrative personnel. The search unit has a commitment of a 24-hour turn-around time. To meet the multi-subject coverage and rapid turn-around expectations of its clientele, the searchers must exploit any search aid or strategy,which would speed up the search process. The storying of hedges is a practical solution to responding to repeated search requests which cannot be retrieved via existing controlled vocabulary, groupings of controlled terms (e.g. MEDLINE's MeSH trees; BRS's pre-cascading of Biosis' biological codes) or a few definitive and specific key-text words. These "popular" requests, though similar, may each have a different focus. The use of a stored or pre-processed hedge as a part of the search strategy, when appropriate, is an obvious cost-effective and time-saving step.

The storing of hedges is a practical solution to responding to repeated search requests which cannot be retrieved via existing controlled vocabulary

The term "hedges" was coined by the Medlars Management Section of the National Library of Medicine for search fragments on selected subjects stored in the computer and accessible by a searcher. "These search fragments—colloquially called "hedges," as opposed to "trees,"—are lists of terms on specific subjects taken from many categories in the tree structures. The terms in a MeSH tree all refer to similar sorts of things—drugs, diseases, etc. In a "hedge" the terms are related by broad subject area, but they cross categories. Thus, for example, a "hedge" on the eye would contain terms dealing with anatomy, diseases, techniques, drugs, diagnosis, etc." (1). The compilation of hedges has evolved to include terms not necessarily part of a controlled vocabulary or hierarchical arrangement. "Hedges may combine terms which are synonyms, variant spellings of the same term, or terms that are related in some other way." (2).

At the NIH Library, searchers are encouraged to store hedges whose construction involves ingenuity or considerable time, or which express a new active research area at NIH. An example of a very useful hedge is the "AIDS" (acquired immune deficiency syndrome) hedge. In the spring of 1981, reports of Kaposi's sarcoma, Pneumocystis carinii pneumonia, and other severe opportunistic infections occurring in homosexual men began to accumulate at the Centers for Disease Control. The Centers for Disease Control also issued a working definition of acquired immune deficiency syndrome (AIDS) to be used in epidemiologic surveillance and clinical diagnosis of cases. According to this definition AIDS is characterized by the occurrence of Kaposi's sarcoma and/or life-threatening opportunistic infections in individuals less than 60 years of age without a known underlying cause for immune deficiency (3,4). The acronym AIDS began to appear in the literature in 1982. Literature requests on AIDS started coming into the NIH Library almost from the beginning of reports on the disease entity. Until the establishment of the MeSH heading for AIDS by MEDLINE during the latter part of 1983, searchers experienced considerable difficulty in designing a search strategy on the subject. The problem was compounded by the nondistinctiveness of the acronym AIDS making it useless for key or text word searching. A search strategy was developed in response to a request for a monthly print-out of the AIDS literature. This strategy, with minor modification, was stored as a separate "save" search, and can be used for the continuing request for retrospective (prior to late 1983) searching of the AIDS literature.

Another example of a very useful and frequently used hedge is the "neuropeptide" hedge. None of the databases has a term or descriptor code or grouping which includes all the individual neuropeptides as a retrievable entity. A hedge based on a table in an article on neuropeptides was created to meet the many requests for searches on various aspects of neuropeptides (5). This hedge is updated as new neuropeptides are identified.

Searchers also try to familiarize themselves with special hedges created by a database producer or by a search service. These pre-processed hedges are created as separate access points to a database and usually are stored under a specific hedge code and unique user number. Examples are: The Excerpta Medica hedges stored by DIALOG and specific to DIALOG. They include such topics as "Adverse Drug Reactions" and "Laboratory Animals." They are listed in the DIALOG Excerpta Medica User Manual. Agricola has stored a number of hedges, such as geographic and nutrition headings. Chemical Abstracts announced a plan to publish "canned search" strategies for various kinds of physical data in its *Online Bulletin*. PsycINFO has stored a number of hedges which can be retrieved on DIALOG. A list of the hedges can be recalled via the ?Psycsave command. They cover such areas as minority groups and developing countries.

The Excerpta Medica hedges incorporate so many descriptor codes and section numbers unique to EM that the application of the hedge is limited to EM. However, hedges stored by Agricola are key words, synonyms, and abbreviations and can be usefully applied to other databases. The hedge "essential amino acids," for example, can be "borrowed" via an execution of the hedge name/user number, "saved," and carried over to another database such as the Nutrition Abstracts and Reviews file of CAB. A strategy for a search request entitled "heavy metal neurotoxicity. . ." was developed using the "heavy metals" hedge from Agricola and appropriate database specific expressions for "neurotoxicity" such as MEDLINE's nervous system diseases/chemically induced, Biosis' concept codes for nervous system pathology, or Excerpta Medica's tag number for "nervous system" and other expressions of neurotoxicity. Using a pre-processed hedge rather than storing a similar one has the additional advantage of avoiding paying monthly storage fees.

A judicious storing of hedges and use of existing and pre-processed hedges then can be a very useful search aid.

1. Pothier, P. E. Hedges, LIBRARY NETWORK/MEDLARS TECHNICAL BULLETIN (January 1976) 9.
2. Dolan, D. R. Hedges for online searching. *DATABASE* 3 (March 1980) 80.
3. Centers for Disease Control. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men – New York City and California. *MMWR* 30 (July 3, 1981) 305-308.
4. Centers for Disease Control. Epidemiologic aspects of the current outbreak of Kaposi's sarcoma and opportunistic infections. *NEW ENGLAND JOURNAL OF MEDICINE* 306 (January 28, 1982) 248-252.
5. Rosenblatt, M. Neuropeptides: future implications for medicine. *MEDICAL TIMES* 111 (November 1983) 32.

Communications to the author should be addressed to Genevieve N. Schiffmann, Technical Information Specialist, National Institutes of Health Library, Ref. & Bibliographic Services Section, 9000 Rockville Pike, Bethesda, MD 20205.

Tailoring Search Strategy for a Highly Specialized Topic. . . .Proof Once Again of the Necessity of Using File Documentation to the Fullest

**by Maureen Lambert
and
Fern Brody
University of Pittsburgh
Library System**

INTRODUCTION

"Quick and dirty" is the searching method most frequently used in a high-volume department which supplies database searches to students and faculty at a large research library. Often, there does not seem to be time to thoroughly review the documentation for unfamiliar files before sitting down at a terminal. Perhaps the patron needs the information immediately, or perhaps the searcher has just completed five searches in a row and would like to move on to other tasks. Yet after having to log off and revise the strategy, or perform costly revisions online, the thought comes to mind that, perhaps it would have been better to do a more thorough review of the documentation.

A 1977 study of strategies of online searching concluded that "little attention was paid to the differences in files and the adaptations in strategy that were necessary because of these differences." (Oldroyd and Citroen, 1977). Tailoring of a strategy may need to be done for several reasons: files can be full-text, bibliographic only or may contain an abstract; the file may lend itself to controlled vocabulary or may be more comprehensively searched free text. We felt that a review of the process of tailoring a search which would span several files might be helpful for both neophytes and experienced searchers who find themselves avoiding the task of wading through a lot of documentation.

An opportunity to perform this exercise presented itself when a music therapist asked for a search. She was interested in the use of music during the childbirth experience, its relationship to relaxation for the mother during labor and delivery, and its ability to reduce the trauma to the newborn. Music therapy was an ideal subject because it encompasses many fields, including medicine, psychology and music. Because therapy techniques are implemented with many different client groups, such as the elderly, deaf or mentally handicapped, it crosses boundaries with disciplines related to those groups.

Terms suggested by our patron included *maternity, obstetrics,* and *midwifery.* We used these terms as a starting point and reviewed documentation for each file. We decided to begin our exercise by using our experience to select the appropriate DIALOG databases.

THE DATABASES

The first database that came to mind to search for this topic was **PsycINFO.** It is a large file and is supported by an excellent thesaurus. Upon consulting this thesaurus and browsing under some suggested related terms we came up with the following straightforward search strategy: **(birth/de or childbirth/de or neonates/de or labor (childbirth)) and music therapy/de.**

Employing our chosen search strategy only yielded one hit. Broadening 'music therapy' to **music/de** netted six more items, most of which were at least of marginal interest to our patron judging from titles such as "The effects of music on the weight loss, crying, and physical movement of newborns".

Judging from our sparse results in PsycINFO, we would not be finding reams of citations on our subject. Our next choice was a file that also covered psychology material, but indexed more medical and general social journals – **MENTAL HEALTH ABSTRACTS.** Until 1983 the producers of MENTAL HEALTH ABSTRACTS had assigned subject headings from a controlled vocabulary list of 120 terms. In 1983 IFI/Plenum Data Company began preparing the file and added 4,000 terms to the controlled vocabulary, but they did not do retrospective indexing. Since we did not want to limit our retrieval to items published in the last two years we searched free text.

Realizing that MENTAL HEALTH ABSTRACTS has lengthy abstracts, we tried to choose broad terms that would not lead us too far astray. Our search strategy looked like this: **(childbirth? or birth? or neonat? or obstetric? or partum or childbearing or parturition or midwife?) and music (w) therapy.** We got four hits, all of which looked good. Next we broadened our search by substituting "music" for "music therapy" and got 25 hits. Of these some were obviously not relevant – "Do the planets govern human personality?" – while some were marginal – "music in general hospital treatment from 1900-1950". Our broader strategy would be appropriate if our patron wanted to do thorough research, rather than a short paper.

The controlled vocabulary for **MEDLINE** is called MESH, short for Medical Subject Headings. Use of this vocabulary is a bit more involved than the usual, and reflects the complexity of the field. There is an alphabetical annotated list of terms which refers one to the permuted terms list by means of codes. The permuted list is arranged by codes, and allows one to see where the term in question falls within a range of broader and narrower headings. For example, the codes for "Music therapy" were **E2.831.440** and **F4.754.549.** (Some terms have ambiguous meanings, which require their placement in two or more areas of medicine.) The **E2** code led us to the heading "Procedure and techniques, therapeutic," while **F4** was "Behavioral tests, therapies, services." In both places, "Music therapy" could be found along with other specialized therapies, such as art or dance. If we had wanted to broaden our search to include these other forms, we could have located more terms by looking at the permuted list. In this case, however, music was what we wanted.

For the other aspect of the search, we were able to make good use of the descriptor codes. Under "Labor" we discovered that there were separate terms for the various stages of labor, with codes beginning **G8.520.769.3.** . . DIALOG enables one to truncate codes, so that we could be sure of not missing any articles which might have been coded specifically for one stage of labor. Alternatively, we could have typed in the term and truncated, i.e. "labor?", but the code enabled us to be precise and yet broad at the same time. Other relevant terms which we decided to include were verified in the thesaurus. Those which consisted of two or more words, the bound descriptors, we entered with their natural language spaces, since according to the documentation, the system would automatically revert to the descriptor field. We entered the single term descriptors with the **/df** qualifier. The final strategy was: **ss music therapy and (DC = G8.520.769? or obstetrics/df or midwifery/df or obstetrical nursing or neonatology/df or perinatology/df or infant, newborn).**

This yielded three postings on file 154 (articles from 1980 to August 1984). Two of the three were of possible interest; the third was irrelevant. We therefore broadened the search to: **music/df** and the birth concepts and got four very good hits.

On the premise that we wanted to be fairly thorough, and additionally because this file contains some psychology materials, we searched **ERIC.** Past experience with this file had led us to view it as very broad in scope. Better results had previously been achieved with thesaurus terms and field delimiters.

Under "Music therapy" in the thesaurus we noted that there were only 28 postings on the CIJE half of the database and 24 on the RIE half. This was a clue that the file would probably yield few hits, but we decided to search it anyway. There were some good terms for the childbirth aspect of the query, including "birth", "obstetrics", and "neonates". Some related (admittedly very peripheral) terms we liked were "perinatal influences" (defined as "factors occurring at the time of birth and affecting the physical or mental development of an individual") and "Infant behavior". Use of the thesaurus also gave us ideas for terms which, though not descriptors in ERIC, suggested approaches to take on other files, i.e. "parturition".

The final strategy used on ERIC was: **ss music (w) therapy and (birth/de or obstetrics/de or perinatal ( ) influences/de or neonates/de or infant (w) behavior/de.** We got zero hits. We then tried linking **music (w) therapy and midwifery/de**, again with no results. Finally we broadened it to **music/de** linked with the childbirth terms **(birth/de, obstetrics/de**, etc.). This strategy yielded 10

▶

postings, most of which had to do with the response of infants to music. Two titles of interest were "The effects of responsive stimulation in early infancy", and "Infant habituation to visual and auditory stimulation". These would probably be of only marginal relevance.

SOCIOLOGICAL ABSTRACTS would be a good source of information from the field of social science and good coverage of conference presentations. Our first thought was to search using controlled vocabulary so we poured over the print verison and selected some terms. The strategy we ran was **(birth/de or childbirth/de or neonat?/de) and music/de**. There was no term for music therapy and we got zero results anyway. Upon consulting the User's Reference Manual we found the suggestion "for the most comprehensive search results, it is recommended that searching not be restricted to the descriptor field" so we also decided to try searching SOCIOLOGICAL ABSTRACTS free text. We used the same strategy that we had employed on Mental Health Abstracts and found six hits, but none were relevant. Just as a test we searched for the word "lamaze". We got three hits but none dealt specifically with our topic.

Both **SOCIAL SCISEARCH** and **SCISEARCH** were chosen because their documentation states that they cover the areas of psychology and psychiatry. Both are produced by the Institute for Scientific Information and are constructed alike. Since there are no descriptors assigned to the articles, the searcher must rely on significant words appearing in the titles. When no field is specified searching is conducted on the basic index, which consists of the article title and the author's addresses. Bonnelly and Drolet (1978) suggest using as many synonyms as possible, including both broad and narrow terms, and using proximity indicators rather than combinations of single terms. Savage and Pemberton (1978) found that the same strategy might be used for both files.

We decided to search SOCIAL SCISEARCH first. The essence of the approach follows: **ss music and (birth? or obstetric?? ? or labor or maternity or neonate? or midwife? or parturition or patrum or perinat?)**. Five of the six articles located were highly relevant, i.e. "The effect of music on relaxation of expectant mothers during labor," and "Music therapy-assisted labor and delivery."

Elated by our success on SOCIAL SCISEARCH, we merely saved the search and executed it on its sister file, SCISEARCH, netting one hit. However, we noticed that their term "music therapy" alone had only one posting, and concluded that it was not as appropriate a file as SOCIAL SCISEARCH for this topic.

DISSERTATION ABSTRACTS ONLINE posed some problems for our search. The file is very big and includes dissertations going back to 1861. Until recently only the contents of Comprehensive Dissertation Index were online, but from July 1980 abstracts began also to be included; searching this file was changed radically. The majority of the records have a small number of words, but the new entries have very wordy abstracts. Another problem was that the database includes both the Humanities/Social Sciences and Science/Engineering sets of Dissertation Abstracts. Though we had to search a broad spectrum of literature for our music therapy topic in this file we were also forced to search such irrelevant disciplines as history, art, mathematics, etc.

We re-used the string of birth concepts first used in MENTAL HEALTH ABSTRACTS anded to "music?" Looking at some of the records revealed that the strategy was too broad. We were probably getting a number of false hits due to the use of "birth" with the word "music". Unfortunately there were only thirty-two occurances of the phrase "music therapy", and none of them mentioned childbirth, partum, etc.

The only exclusively music database available on DIALOG, RILM, was searched using a combination of thesaurus and free text terms, along with some codes. The only relevant term in the thesaurus was the general term "Therapy"; the codes included one for "Psychology, therapy and medicine" and "Psychology and hearing". We decided to try a simple strategy first: **ss therapy/de and (birth) or obstetric?? ? or labor? or maternity or neonat?? ? or midwifery or labour)**. Since RILM is an international database, we decided to included the British spelling; although titles for non-English works are given in English, it was not clear whether American or British spellings were preferred.

The result was one hit, which was relevant, but in Polish: "Music therapy in Gdansk: its use in gynecology and midwifery." To broaden our approach we tried combining the codes mentioned above with the free-text terms used in the first try. None of the six articles we found were relevant. Since we had obtained such poor results, we decided to double-check by combining the broad term "music" with the codes used previously and turned up only nine articles of general music therapy interest but not of relevance to this topic.

CONCLUSIONS

We felt that the exercise was useful from two angles: we learned which files were most fruitful for searching for music therapy material, and we relearned how to tailor our strategies to suit the files. Being consistent in terminology across files did not aid our retrieval. Flexibility was important in our approach. On more difficult files like Medline and PsycINFO it helped to have back-up or alternative plans in case one search query failed. Occasionally we found it helpful to re-check our findings in various files by searching for a term that was related to our topic—for example "Lamaze" and the codes in RILM. The records that were pulled up with this strategy could have supplied overlooked terms to use in the search.

> **Being consistent in terminology across files did not aid our retrieval. Flexibility was important in our approach. On more difficult files like Medline and PsycINFO it helped to have back-up or alternative plans in case one search query failed.**

Use of the various thesauri helped us to compile lists of terms for free-text searching. In general it might be better to consult as many subject heading lists as possible while setting up the strategies. In an inter-disciplinary search such as this, more technical terms must be combined with the general because of the differing viewpoints presented in the various disciplines. Here again use of the thesauri helps to key in on jargon, along with reliance on the patron's knowledge.

In almost every case, few hits caused us to re-think and re-do our strategy to try to pull in as many marginally interesting citations as possible so that our patron would have some place to start her research. We found that our exercise does confirm the adage about best laid plans falling awry.

BIBLIOGRAPHY

Angier, Jennifer J., "Online Searching in Mental Health: the National Institute of Mental Health Database", *Database* (v. 4 #4), December 1981, p 73-78.

Bonnelly, Claude and Gaetan Drolet, "Searching the Social Sciences Literature Online: Social SciSearch." *Database* (v. 1 #4) December 1978, p 10-25.

DIALOG Chapters. DIALOG Information Services, Inc.

Epstein, Barbara A. and Jennifer J. Angier, "Multi-database Searching in the Behavioral Sciences. Part 1: Basic Techniques and Core Databases", *Database* (v. 3 #3), September 1980, p 9-15.

Hawkins, Donald T, "Multiple Database Searching: Techniques and Pitfalls", *Online* (v. 2 #2), April 1978, p 9-15.

Oldroyd, Betty K. and Charles L. Citroen, "Study of strategies used in on-line searching." *On-Line Review* (v. 1 #4), p 295-310.

Pugh W. Jean, "A Bibliography of Database and Search System Comparisons", *Online* (v. 6 #5), September 1982, p 41-55.

Savage, Gretchen S. and Jeff Pemberton, "Scisearch on Dialog." *Database* (v. 1 #3), September 1978, p 50-67.

Sociological Abstracts User's Reference Manual, prepared by Miriam Chall, revised edition, San Diego: Sociological Abstracts, 1983.

ACKNOWLEDGEMENTS

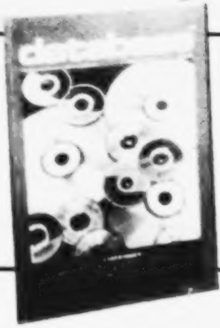
Special thanks to Sr. Donna Marie Beck, C.S.J., Director of Music Therapy at Duquesne University, Pittsburgh, Pa., and to Professor Carolyn Nickerson of the St. Francis College School of Nursing, Loretto, Pa., who provided the subject expertise.

THE AUTHOR

MAUREEN LAMBERT received her Advanced Certificate in Library Science in 1983 from the School of Library and Information Sciences at the University of Pittsburgh. She presently performs reference and database searching at Pitt's Graduate School of Business.

FERN BRODY is Coordinator of Database Services at the University of Pittsburgh.

Communications to the authors should be addressed to Maureen Lambert and/or Fern Brody, 136 Mervis Hall, University of Pittsburgh, PA. 15260.



THE
"SEARCHER'S
JOURNAL". . .

For Professional Database Searchers. . .
And Database Builders

database®

SIX KEY REASONS

Why Thousands Of Searchers Keep *DATABASE* Near Their Terminal

1. In-depth Articles Keep You Up-To-Date On New Databases In **Business, Science And Technology, Life Sciences, Law, Arts** And **Humanities**

2. Current Coverage Of Databases On **CDROMs And Other Laserdisks**. Reviews And Case Histories.

3. **Microcomputers In Libraries**. A Steady Flow Of Software Reviews, Hardware Evaluations, Applications Ideas, And Program Codes.

4. Database Creation . . . **How-to Articles On Building Your Own File** . . . On Magnetic Disks Or CDROM Or WORM.

5. **News, Special Columns**, Book Reviews, Editorials And Letters.

6. Lots Of **Charts, Lists, Bibliographies**, Directories And Search Printouts.

Consider The Useful, Money-Saving Tips In This Sampling Of Recent Articles. . .

"Establishing A Business Intelligence Clearinghouse". . . "The Beilstein Chemistry Database". . . "Sources of Japanese Business Information". . . "CDROM vs. Online vs. Data Tapes vs. Print". . . "Searching Current Events". . . "Crossfile Searching For Patents". . . "Online Sources of Competitive Intelligence". . . "Online Tax Information". . . "Laserdisk Directory" (five part series). . . "PROMT – The Megafile". . . "Guide to DBMS Software". . . "Full Text Searching". . . "Review of STAR Software". . . "Regional Business Information". . . "LEXIS and WESTLAW Downloading Techniques". . . "Comparison of Low Cost Communications Software for PCs". . . "INDEXIT Software Review". . . "INMAGIC . . . Kudos and Caveats". . . "Survey of Computer Databases". . . "Online and Offline Cost Comparison". . .

***DATABASE* Will More Than Pay Its Cost By Helping Your Search Proficiency. . .And We Offer A No-risk, Money Back Guarantee To Prove It. . . Plus 40% Discounts On Personal, Multiple Organizational And Small Public/Academic Library Orders**

**Order Now Toll-Free – Or Use The Coupon Below
800/248-8466**

To: Online, Inc., 11 Tannery La., Weston, CT 06883

Yes, I'll try *DATABASE* on a no-risk basis. If not satisfied I can cancel and keep the first issue free.

☐ *DATABASE* (6X/yr) organizational price: $85
☐ *DATABASE* 40% discount (personal checks, home delivery. . .second & additional organiz. orders to same address. . .pub./academ. libraries under $100K/yr materials budget: $51
Mexico and C.A., please add $15/yr.
All other international (airmail only), add $35/yr.

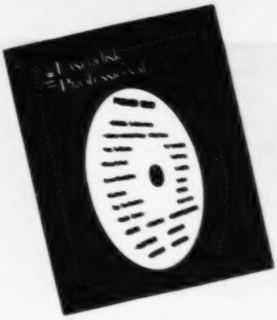
Please include check or credit card number with order.

Send to ______________________
Company ______________________
Address ______________________
City ______________________
State ______________ Zip ______________

Charge to ☐ AMEX ☐ VISA ☐ MasterCard

Card no. ______________________
Expiration date ______________________

ANNOUNCING THE LASERDISK PROFESSIONAL™



AT LAST. . .A Practical, How-to Journal For Laserdisk Librarians With The Editorial Excellence of ONLINE And database

The Laserdisk Professional Will Bring You A Full Range Of Original Articles And Columns. . .Some Examples:

ARTICLES. . .

- **CDROMs As Online Hosts** On Local Networks.
- **Laserdisks In Libraries**. . .Who Is Using Them And What Do They Like—Or Dislike? The First Major Survey In The Field.
- **Public Access** To Laserdisks. . .Experts' Reports On Problems And Solutions.
- **How To Buy** Your First CDROM Workstation. . .Do's And Don'ts.
- **The Seven Deadly Sins** Of CDROM Usage. . . And How To Avoid Them.
- **The Top Ten** Laserdisks Bestsellers—By Library Type.
- **Future Tech**. . .Some Likely Scenarios Beyond CDROMs.

PLUS. . .

- **Product PREviews**. . .Reports From The Beta Test Sites
- **Individual Product** Reviews. . .User Evaluations Of Commercially Available Laserdisks
- **Product Comparisons**. . .Feature-by-Feature Analysis Of Disks With Similar Databases
- **NEWS**. . .New Databases On Disk, New Hardware. . .A Roundup Of Developments In The Full Range Of Optical Disk Technology
- **Regular Columns**. . ."Hardware Highlights"—Features And Photos Of Newly Available Equipment. . .Plus "Laserdisk Library Profiles"—Case Histories Of Successful Installations
- **Readers' Questions**. . .Experts Answer Your Questions About Hardware And Database Selection, Installation, Patron Service, Funding And Marketing
- **Book & Magazine Reviews**. . .Evaluations Of Currently Available Publications
- **Letters** To The Editor, **Interviews** With Pioneering Librarians And Industry Executives, And Thoughtful Editorial Comment

Premier issue:
April, 1988
Frequency: bimonthly
Price: $78/yr.
Editor: Nancy K. Herther
Publisher: Adam C. Pemberton
Order Now To Be Sure To Get The Premier Issue
Call Toll-free (800) 248-8466. . .Or Use The Attached Coupon

ONLINE THE Laserdisk Professional database

Special Offer For Prepaid Subscribers. . .A Free IBM Compat. Demo Floppy Of DIALOG OnDisc CDROM. . .

Mail To: Laserdisk Professional
11 Tannery Lane, Weston, CT 06883

☐ Yes, I'd like to start the *Laserdisk Professional* with the Premier Issue and receive the free offer.
☐ The *Laserdisk Professional* (6X/yr.)—$78/yr. U.S.
☐ Mexico and C.A. please add $15/yr.
☐ All other international, add $30 (airmail only)

Name ____________________
Organization ____________________
Street ____________________
City ____________ State/Province ____________ Zip ____________
Country ____________________
☐ AMEX ☐ VISA ☐ MasterCard
Acct # ____________ Exp. date ____________

Please enclose check to receive free offer. If not satisfied, you may cancel and keep first issue and gift.
Or ORDER TOLL-FREE 800/248-8466 with AMEX, VISA or MasterCard.